[illegible]

SANKHYA: THE INDIAN JOURNAL OF STATISTICS

CONTENTS OF VOLUME SEVEN *1945-46*

(*All Rights Reserved*)

PAPERS

Pages

Editorial Note 1
On Grassmann and Differential Rings and their relations to the Theory of Multiple Integrals. *By S. S. Chern.* 2
On the Generalisation of Markoff's Theorem and Tests of Linear Hypothesis. *By C. R. Rao* 9
Markoff's Theorem with Linear Restrictions on Parameters. *By C. R. Rao* .. 16
On the Accuracy of the Different Approximations to the L_1-Distribution. *By Miss. A. George* 20
A Note on the Distribution of Chi-Squares. *By A. Bhattacharyya.* 27
Report on the Bihar Crop Survey : Rabi Season 1943-44. *By P. C. Mahalanobis* .. 29
The Logical Inversion of the Notion of the Random Variable. *By R. A. Fisher* .. 129
The Individual Sampling Distributions of the Maximum, the Minimum and any Intermediate of the p-Statistics on the Null-Hypothesis. *By S. N. Roy* .. 133
Distribution of Certain Frequency Constants in Samples from Non-normal Populations. *By K. C. Cheriyan* 159
Statistics and Social Values. *By N. C. Mehta* 167
Statistics in the Service of Planning. *By R. K. Mukherjee* 173
Differential Rates of Natural Increase. *By N. G. D. Joarder* 176
Cost of Living Indices at Cawnpore. *By S. P. Saksena* 179
The Cawnpore Mill-worker, His Family and His Expenditure. *By J. K. Pande* .. 185
Agricultural Co-operation in Bengal and Rest of British India (1918-19 to 1938-39). *By T. Ghosh and H. Sinha* 189
On the Totality of the Solutions for the Symmetrical Incomplete Block Designs : $\lambda=2$, $k=5$ or 6. *By Q. M. Husain* 204
On the Linear Combination of Observations and the General Theory of Least Squares. *By C. R. Rao* 237
Theories of Probability. *By J. G. Singh* 257
A Statistical Approach to the Problem of Chronology of Shakespear's Plays. *By M. R. Yardi* 263
Sample Surveys of Crop Yield in India. *By P. C. Mahalanobis* 269
Sampling for Rice Yield in Bihar and Orissa. *By J. A. Hubback* 281
Confounded Designs in Quasi-Latin Squares. *By C. R. Rao* 295
Enumeration of Non-Isomorphic Solutions of Balanced Incomplete Block Designs. *By H. K. Nandi* 305

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

SANKHYĀ: THE INDIAN JOURNAL OF STATISTICS—VOLUME SEVEN

Pages

A Further Note on Non-Isomorphic Solutions of Incomplete Block Designs. *By H. K. Nandi* 313

Impossibility of Symmetrical Incomplete Block Design with λ=2, k=7. *By Q. M. Husain* 317

A Sample Survey of After-Effects of the Bengal Famine of 1943. *By P. C. Mahalanobis, R. K. Mukherjea and A. Ghosh* 337

On a Measure of Divergence of two Multinomial Populations. *By A. Bhattacharyya* .. 401

Tests with Discriminant Functions in Multivaria'e Analysis. *By C. R. Rao* .. 407

Confidence Interval for the Correlation Coefficient. *By K. C. S. Pillai* 415

A New Solution in Symmetrical Balanced Incomplete Block Designs. *By K. N. Bhattacharyya* 423

A Study of Expenditure Pattern of Calcutta Middle Class Family Budgets. *By S. J. Poti* 425

Distribution of Muslims in the Population of India. *By P. C. Mahalanobis* .. 429

MISCELLANEOUS NOTES

Bernoullis Theorem and Tshebyscheff's Analogue .. *S. N. Roy and P. K. Bose* .. 209

Incidence of Ticketless Travelling on the Lucknow Cawnpore Section of the E. I. Railway .... *J. Singh* 211

Influence of Humidity and Temperature on the Yield of Cotton .. .. .. *K. R. Nair and P. K. Bose* .. 213

On the Order of Approximation Involved in Laplace's Central Limit Theorem in probability *M. N. Ghosh* 323

A Note on Concentration of Income in Bengal villages .. .. .. *R. K. Mukherjea and M. M. Mukherjee* 327

A Criterion for Testing whether Two Samples have come from the Same Population without assuming the Nature of the Population .. *K. K. Mathen* 329

Price of Agricultural Commodities and Land Transfer in Bengal .. .. .. *N. R. Das Gupta* 330

Test of Significance for Intra-class Correlation when the Family Sizes are not equal .. *R. P. Bhargava* 435

On Locally Most Powerful Tests when Alternatives are Onesided .. .. *C. R. Rao and S. J. Poti* .. 439

On the Most Efficient Designs in Weighing. .. *C. R. Rao* 440

A Note on Concordance .. .. *M. Mukherjee* 441

A Note on Concentration of Agricultural Wealth in Bengal .. .. .. *R. K. Mukherjea* 442

A Statistical Study of the Labour in the Assam Tea Plantation .. .. .. *M. V. Seshagiri Rao* .. 445

REVIEWS

		Pages
The Population Problem in India	*N. T. Mathew*	331
Student's Collected Papers	*K. R. N*	331
Tables of Powers	*P. K. Bose*	332
A Revew of Price Control in India	*T. Choudhury*	332
Savings and Investments in India	*Durgadas Bose*	333
Conflicts of Economic Ideologies in India	*M. Mukherjea*	333
Examination in India	*C. Das Gupta*	333
Building the Canadian West	*P. C. Bandopadhyay*	334
The Drama of Weather	*P. C. Bandopadhyay*	335
The Forming of Federations	*C. Das Gupta*	336

REPORTS

Indian Statistical Institute : Twelfth Annual Report : 1943—44 107
Thirteenth Annual Report : 1944—45 217

AUTHOR INDEX

Bhattacharyya, A. A Note on the Distribution of Chi-Squares 27
———— On a Measure of Divergence of two Multinomial Populations. .. 401
Bhattacharyya, K. N. A New Solution in Symmetrical Incomplete Block Designs .. 423
Bhargava, R. P. Test of Significance for Intra-class Correlation when Family Sizes are not Equal 435
Cheriyan, K. C. Distribution of Certain Frequency Constants in Samples from Non-normal Populations. 159
Chern, S. S. On Grassmann and Differential Rings and their Relations to the Theory of Multiple Integrals 2
Dasgupta, N. R. Price of Agricultural Commodities and Land Transfer in Bengal .. 330
Fisher, R. A. The Logical Inversion of the Notion of the Random Variable 129
George, Miss. A. On the Accuracy of the Different Approximations to the L_1-Distribution 20
Ghosh, M. N. On the Order of Approximation Involved in Laplace's Central Limit Theorem in Probability 323
Ghosh, T and Sinha, H. Agricultural Co-operation in Bengal and Rest of British India (1918-19 to 1938-39) 189
Hubback, J. A. Sampling for Rice Yield in Bihar and Orissa 281
Husain, Q. M. On the Totality of the Solutions for the Symmetrical Incomplete Block Designs : $\lambda=2$, $k=5$ or 6 204
———— Impossibility of Symmetrical Incomplete Block Designs with $\lambda=2, k=7$ 317
Joarder, N. G. D. Differential Rates of Natural Increase 176
Mahalanobis, P. C. Report on the Bihar Crop Survey :Rabi Season 1943-44 .. 29
———— Sample Surveys of Crop Yield in India 269
———— Distribution of Muslims in the Population of India .. 429
Mahalanobis, P. C., R. K. Mukherjea and A. Ghosh. A Sample Survey of After-Effects of the Bengal Famine of 1943. 337

Pages

Mathen, K. K. A Criterion for Testing whether two Samples have come from the Same Populations without assuming the Nature of the Population. .. 329

Mehta, N. C. Statistics and Social Values 167

Mukherjee, R. K. Statistics in the Service of Planning 173

Mukherjea, R. K. A Note on Concentration of Agricultural Wealth in Bengal .. 442

Mukherjee, R. K. and Mukherjee, M. M. A Note on Concentration of Income in Bengal Villages 327

Mukherjee, M. M. A Note on Concordance 441

Nandi, H. K. Enumeration of Non-Isomorphic Solutions of Balanced Incomplete Block Designs.. 305

——— A Further Note on Non-Isomorphic Solutions of Incomplete Block Block Designs 313

Nair, K. R. and Bose, P. K. Influence of Humidity and Temperature on the Yield of Cotton 213

Pillai, K. C. S. Confidence Interval for the Correlation Coefficient 415

Pande, J. K. The Cawnpore Mill-worker, His Family and His Expenditure .. 185

Poti, S. J. A Study of Expenditure Pattern of Calcutta Middle Class Family Budgets 425

Rao, C. R. On the Generalisation of Markoff's Theorem and Tests of Linear Hypotheses 9

——— Markoff's Theorem with Linear Restrictions on Parameters .. 16

——— On the Linear Combination of Observations and the General Theory of Least Squares 237

——— Confounded Designs in Quasi-Latin Squares 295

——— Tests with Discriminant Functions in Multivariate Analysis .. 407

——— On the most Efficient Designs in Weighing 440

Rao, C. R. and Poti, S. J. On locally Most Powerful Tests when Alternatives are Onesided 439

Rao, M. V. S. A Statistical Study of the Labour in the Assam Tea Plantation .. 445

Singh, J. G. Theories of Probability 257

——— Incidence of Ticketless Travelling on the Lucknow Cawnpore Section of the East Indian Railway 211

Roy, S. N. The Individual Sampling Distributions of the Maximum, the Minimum and any Intermediate of the p-Statistics on the Null-Hypothesis .. 133

Roy, S. N. and Bose, P. K. Bernuolli's Theorem and Tshebyscheff's Analogue .. 209

Saksena, S. P. Cost of Living Indices at Cawnpore 179

Yardi, M. R. A Statistical Approach to the Problem of Chronology of Shakespeare's Plays 263

SANKHYĀ
THE INDIAN JOURNAL OF STATISTICS

Edited By P. C. MAHALANOBIS

VOL. 7 PART 1 1945

EDITORIAL NOTE

We offer our apologies to the members of the Indian Statistical Institute and subscribers to Sankhyā for the delay in publishing Vol. 7. Early in 1942 a big portion of the Statistical Laboratory including the main library and all valuable records had to be hurriedly evacuated to Giridih, a place 206 miles away from Calcutta. Owing to shortage of paper, lack of printing facilities, and difficulties arising from the war situation in north east India we were unable for a long time to arrange regular publication of Sankhyā. Printing of the current number was further delayed by the promulgation of the Paper Control Order by the Government of India in 1944 which required the number of pages of every journal being reduced by 70 per cent. As a result of our representations special sanction was given by the Government of India relaxing the above limit to a considerable extent in our case. Some of the other difficulties have also been overcome and we hope to be able to issue the journal more regularly in future.

Owing to the dislocation in our printing and publishing arrangements a large number of papers have accumulated for which we offer our apologies to the contributors. We intend to print all accepted papers as quickly as possible but in many cases we shall have to do this in an abstract form. We shall also be obliged to publish all articles in as condensed a form as possible until paper position improves, and we earnestly seek the co-operation of all our contributors in this matter.

P. C. MAHALANOBIS

ON GRASSMANN AND DIFFERENTIAL RINGS AND THEIR RELATIONS TO THE THEORY OF MULTIPLE INTEGRALS

By SHIING-SHEN CHERN
Tsing Hua University, Kunming, China.

INTRODUCTION

During my lectures* given under a friendly invitation of the University of Calcutta I have spoken about the theory of Grassmann and differential rings. The role that this calculus plays in the theory of multiple integrals is well-known and it is of little doubt that it will be a useful tool in many statistical problems. I compile in this paper the fundamental propositions of the theory. The paper being expository in character, I have freely drawn upon results from the sources which I give in the bibliography at the end of the paper.

2. AXIOMS AND FUNDAMENTAL PROPOSITIONS OF A GRASSMANN RING

Let R be a ring with unity element and consider an n-dimensional vector space G over R (Van der Waerden : 1937). Let $x_1, x_2, \ldots, x_n$ be a set of basis elements of G. To define a Grassmann ring we suppose the products of basis elements be defined, with the properties :

(1) The multiplication is associative : $x_i\,(x_j x_k) = (x_i x_j).x_k$;

(2) $x_i x_j = -x_j x_i$, (3) $x_i^2 = 0$.

We take the unity element of R (denoted by 1) and the different products $x_{i_1} \ldots . x_{i_r}$, $(r=1, \ldots, n)$, as the basis elements of a vector space of 2^n dimensions over R and denote this vector space by R $(x_1, \ldots, x_n)$ or simply R(x). By supposing the two distributive laws to hold :

$$a\,(b+c) = ab+ac, \quad (b+c)\,a = ba+ca, \quad a,b,c \subset \mathrm{R}(x),$$

we can define a multiplication in R$(\,x\,)$. Then R(x) becomes a ring and is a hyper-complex system or an algebra of rank 2^n over the ring R. Owing to this particular way of definition R(x) is called a Grassmann ring or a Grassmann algebra of dimensions n over R.

We proceed to derive some fundamental propositions of a Grassmann ring.

First of all, it is evident that an element of R(x) can be written in one and only one way in the form

$$x = x^{(o)} + x^{(1)} + \ldots + x^{(n)} \qquad \ldots (2)$$

where $x^{(o)} \varepsilon \mathrm{R}$ and

$$x^{(r)} = \Sigma \alpha_{i_1 \ldots i_r}\, x_{i_1} \cdots x_{i_r}, \quad \alpha_{i_1 \ldots i_r}\, \varepsilon \mathrm{R}, \; r=1,\ldots,n \qquad \ldots (2{\cdot}1)$$

the summation being extended over all combinations of 1, ... 2,, n taken r at a time, with $\mathrm{i}_1 < \ldots < \mathrm{i}_r$. We can also write

$$x^{(r)} = \frac{1}{r!} \sum_{i_1,\ldots,i_r=1}^{n} \alpha_{i_1 \ldots i_r} x_{i_1} \cdots x_{i_r} \qquad \ldots (2{\cdot}2)$$

---

* Calcutta, July 21, 22, 23 and 27, 1943.

if we assume the coefficients $\alpha_{i_1 \dots i_r}$ to be skew-symmetric in the lower indices. In both expressions (2·1) and (2·2) we have $x^{(r)}=0$ when and only when $\alpha_{i_1 \dots i_r}=0$. In analogy with the case of a polynomial ring x is sometimes called a polynomial. This polynomial is said to be of degree r if

$$x^{(r)} \neq 0, \quad x^{(r+1)} = \dots = x^{(n)} = 0.$$

If $x = x^{(r)}$, we say that it is a form of degree r. An element $x = \alpha x_{i_1} \dots x_{i_r}$ is called a monomial.

Theorem 1. If x and y are two forms of degrees r and s respectively, then $yx = (-1)^{rs} xy$.

Proof. If x and y are monomials, the theorem is evident, for we can then move the factors of y to the other side of x, causing altogether rs changes of sign. The general formula then follows from the distributive laws.

In a Grassmann ring the notion of linear dependence takes a particularly simple form. To establish this notion we shall suppose R to be a field.

Let $u_1, \dots, u_r$ be r elements of $R(x_1, \dots, x_n)$. They are called linearly dependent, if there exist $\alpha_1, \dots, \alpha_r \subset R$, not all zero, such that $\alpha_1 u_1 + \dots + \alpha_r u_r = 0$. If the equation holds only when $\alpha_1 = \dots = \alpha_r = 0$, the elements $u_1, \dots, u_r$ are called linearly independent. Of particular importance is the case in which $u_1, \dots, u_r$ are linear forms. Then we have :

Theorem 2. The linear forms $u_1, \dots, u_r$ are linearly dependent, if and only if $u_1 \dots u_r = 0$.

Proof. To prove this let us put $u_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i=1, 2, \dots, r.$

Then we have

$$u_i \dots u_r = \sum_{i_1 < \dots < i_r} \begin{vmatrix} a_{1i_1} \dots a_{1i_r} \\ \dots \\ a_{ri_1} \dots a_{ri_r} \end{vmatrix} x_{i_1} \dots x_{i_r}.$$

where the summation is extended over all combinations of 1, 2, …, n. From this relation the theorem follows.

Corollary. If $u_1 \dots u_{r-1} \neq 0$, $u_1 \dots u_r = 0$, then u_r is a linear combination of $u_1, \dots, u_{r-1}$

3. AN EXAMPLE

The object of this section is to give an example of a Grassmann ring. Historically this example was essentially the way that Grassmann first formulated this calculus. We give here this example, not only because of its historical and geometrical interest, but also of the fact that it proves the consistency of our axioms.

Consider the ring R and adjoin to it n sets of variables

$$x_1^{(1)}, \dots, x_n^{(1)}; \; x_1^{(2)}, \dots, x_n^{(2)}; \; \dots; \; x_1^{(n)}, \dots, x_n^{(n)}.$$

We put

$$[x_i] = x_i^{(j)}, \quad i = 1, \dots, n,$$

and $$[x_{i_1} \dots x_{i_r}] = \begin{vmatrix} x_{i_1}^{(j_1)} & \dots & x_{i_r}^{(j_r)} \\ & \dots & \\ x_{i_1}^{(j_r)} & \dots & x_{i_r}^{(j_r)} \end{vmatrix}; \quad i_1, \dots i_r, j_1, \dots, j_r = 1, \dots, n, \; r \leqslant n.$$

To define a Grassmann ring we consider expressions of the form

$$f\left(x_1^{(j_1)}, \dots, x_n^{(j_1)}; \dots; \; x_1^{(j_r)}, \dots, x_n^{(j_r)}\right),$$

which have the following properties : (1) they are linear in each set of variables ; (2) they change their sign when two sets of variables are interchanged. As elements of our Grassmann ring we then take the sums of such expressions and elements of R. Denoting their totality by R(x), we see that an element of P(x) is of the form

$$x = \alpha + \Sigma \alpha_i [x_i] + \dots + \sum_{i_1 < \dots < {}_r} \alpha_{i_1 \dots i_r} [x_{i_1} \dots x_{i_r}] + \dots,$$

where the α's belong to R. In R (x) there are defined addition and multiplication by an element of R. The multiplication of two arbitrary elements of R(x) is defined by the rules that the associative and distributive laws are universally valid and that the product of two "bracketed expressions" is defined by

$$\left[x_{i_1} \dots x_{i_r}\right]\left[x_{j_1} \dots x_{j_s}\right] = \left[x_{i_1} \dots x_{i_r} x_{j_1} \dots x_{j_s}\right], \quad \text{if } r+s \leqslant n \quad \dots \quad (3{\cdot}1)$$

$$\left[x_{i_1} \dots x_{i_r}\right]\left[x_{j_1} \dots x_{j_s}\right] = 0, \quad \text{if } r+s > n. \quad \dots \quad (3{\cdot}2)$$

With addition and multiplication so defined it is not difficult to verify that the elements of R(x) form a Grassmann ring. The multiplication in the ring is called exterior multiplication, in order to distinguish it from ordinary multiplication.

4. DIFFERENTIAL RINGS

From a Grassmann ring we can arrive at the notion of a differential ring by introducing an operation, called differentiation, concerning which we take as first axiom the following :

1. There exists a ring R$'$, containing R, such that to every element Rε α there corresponds in R$'(x_1, \dots, x_n)$ a unique form

$$\alpha' = \sum_{i=1}^{n} \delta_i \alpha \, x_i, \quad \dots \quad (4)$$

which is zero or linear, called the differential of α. The differentials of sums and products in R satisfy the equations

$$(\alpha + \beta)' = \alpha' + \beta', \quad \dots \quad (4{\cdot}1)$$

$$(\alpha\beta)' = \alpha\beta' + \beta\alpha' \quad \dots \quad (4{\cdot}2)$$

If this axiom is satisfied , we say that R is differentiable in R$'$.

We extend the definition of differential to R $(x_1, \ldots, x_n)$ by supposing the following property to hold

$$(x+y)'=x'+y', \quad x,y \subset R(x_1, \ldots, x_n). \qquad (4{\cdot}21)$$

and by defining the differential of a monomial to be

$$x=\alpha x_{i_1} \ldots x_{i_r}.$$

$$x'=\alpha' x_{i_1} \ldots x_{i_r}.$$

It follows that the differential of a polynomial of degree r in $R(x)$ is at most of degree $(r+1)$.

For the differential of the product of two forms we prove easily the theorem : If x and y are forms of degrees r and s respectively, then

$$(xy)'=x'y+(-1)^r xy'. \qquad (4{\cdot}3)$$

In fact, when x and y are monomials, the relation follows immediately from the definition. The general formula then follows from the distributive law and the relation (4.21).

With a law of differentiation defined in the Grassmann ring the ring will be called a differential ring, if, besides the last axiom, the following axiom is satisfied :

$$(\alpha')' = 0.$$

2. The ring R′ is differentiable in a ring R″ and, if $\alpha \varepsilon R$, we have

$$(\alpha')' = \sum_{i=1}^{n} (\delta_i \alpha)' x_i = \sum_{i,j=1}^{n} \delta_j\, (\delta_i \alpha) x_j x_i.$$

Hence the second axiom is equivalent to the condition $\delta_j(\delta_i\alpha)-\delta_i(\delta_j\alpha)=0$. From this axiom follows the important theorem given below.

Theorem 3 (Poincaré). For any $x \varepsilon R(x)$ we have $(x')'=0$.

Proof. Because of (4·21) it is clearly only necessary to prove the theorem for the case that x is a monomial :

$$x=\alpha x_{i_1} \ldots x_{i_r}.$$

In this case we have

$$(x')'=\sum_{i,j=1}^{n} \delta_j(\delta_i\alpha) x_j x_i x_{i_1} \ldots x_{i_r}=0.$$

Hence the theorem.

The most important example of a differential ring is constructed as follows : we take R to be a ring of functions $f(u_1, \ldots, u_n)$ of n real variables $u_1, \ldots, u_n$, which are differentiable up to the second order inclusive. We construct over R the Grassmann ring $R(du_1, \ldots, du_n)$ or $R(du)$, whose basis elements are the differentials $du_1, \ldots, du_n$. The Grassmann ring $R(du)$ will become a differential ring, if we define

$$f'=df=\sum_{i=1}^{n} \frac{\delta f}{\delta u_i}\, du_i. \qquad (4{\cdot}4)$$

A form in the differential ring $R(du)$ is called an exterior differential form.

5. TRANSFORMATION FORMULAE OF MULTIPLE INTEGRALS

Let Ω be an exterior differential form of degree r :

$$\Omega = \frac{1}{r!} \sum_{i_1,\dots,i_r=1}^{n} \alpha_{i_1 \dots i_r}\, du_{i_1} \cdots du_{i_r}, \qquad \dots \quad (5)$$

where $\alpha_{i_1 \dots i_r}$ are functions of $u_1, \dots, u_n$ and are skew-symmetric in their indices.

We shall define the value of the integral $\int_D \Omega$ over an oriented simplicial domain D of an n-dimensional manifold M. By such a domain we shall mean a domain which is a sum of coherently oriented simplexes in a simplicial decomposition of M. Symbolically we can write,

$$D = \sum_{i=1}^{\mu} \lambda_i \overset{i}{E}_r$$

where $\overset{i}{E}_r$ are oriented r-dimensional simplexes and $\lambda_i = 0, 1, -1$. We then define

$$\int_D \Omega = \sum_{i=1}^{\mu} \lambda_i \int_{\overset{i}{E}_r} \Omega. \qquad \dots \quad (5{\cdot}1)$$

From this definition it follows that the value of the integral is invariant under subdivisions of D and changes its sign when its orientation is reversed.

To give a definition of the integral in question we need only define its value over an oriented simplex, say E_r. Let E_r be defined by the equations

$$u_i = u_i(t_1, \dots, t_r), \quad i = 1, \dots, n. \qquad \dots \quad (5{\cdot}2)$$

Then we define

$$\int_{E_r} \Omega = \frac{1}{r!} \int_{E'_r} \sum_{i_1,\dots,i_r=1}^{n} \alpha_{i_1 \dots i_r} \frac{\delta(u_{i_1}, \dots, u_{i_r})}{\delta(t_1, \dots, t_r)}\, dt_1 \cdots dt_r$$

where E'_r denotes the domain of the space $(t_1, \dots, t_r)$ whose image under (5.2) is E_r. The last integral, being an integral of degree r over a domain in an r-dimensional manifold, is defined in textbooks on analysis.

This definition is independent of the choice of E'_r and the coordinates $t_1, \dots, t_r$. In fact, let

$$t_i = t_i(t_1^*, \dots, t_r^*), \; i = 1, \dots, r,$$

be a one-one transformation of the domain E''_r of the space $(t_1^*, \dots, t_r^*)$ on to E'_r.

Then we have

$$\sum \alpha_{i_1 \dots i_r} \frac{\delta(u_{i_1}, \dots, u_{i_r})}{\delta(t_1, \dots, t_r)}\, dt_1 \dots dt_r = \sum \alpha_{i_1 \dots i_r} \frac{\delta(u_{i_1}, \dots, u_{i_r})}{\delta(t_1, \dots, t_r)} \frac{\delta(t_1, \dots, t_r)}{\delta(t_1^*, \dots, t_r^*)}\, dt_1^* \dots dt_r^*$$

$$= \sum \alpha_{i_1 \dots i_r} \frac{\delta(u_{i_1}, \dots, u_{i_r})}{\delta(t_1^*, \dots, t_r^*)}\, dt_1^* \dots dt_r^*$$

which shows that the value of the integral depends only on E_r.

The fruitfulness of the notion of an exterior differential form in the theory of multiple integrals is based on the following two theorems.

Theorem 4. The value of the integral $\int_D \Omega$ is independent of the choice of the coordinates $u_1, \ldots, u_n$ on the underlying manifold M.

Proof. Let a change of coordinates on M be defined by the equations

$$u_i = u_i(\overset{*}{u}_1, \ldots, \overset{*}{u}_n), \quad i = 1, \ldots, n$$

With the new coordinates $\overset{*}{u}_1, \ldots, \overset{*}{u}_n$ we have

$$\Omega = \frac{1}{r!} \sum \alpha_{i_1 \ldots i_r} \frac{\delta(u_{i_1}, \ldots, u_{i_r})}{\delta(u_{j_1}{}^*, \ldots u_{j_r}{}^*)} \, du_{j_1}{}^* \ldots du_{j_r}{}^*,$$

the summation being extended to all combinations of $i_1 < \ldots < i_r$ and to all $j_1, \ldots, j_r = 1, \ldots n$, That the value of the integral remains unchanged can be immediately verified from the definition.

The ordinary definition of a multiple integral depends on the choice of coordinates so that a theorem is necessary which gives the transformation of a multiple integral under a change of coordinates. By making use of the notion of an exterior differential form, this theorem takes a patricularly simple form.

Theorem 5 (Stokes). Let C be a domain whose boundary is D. Then $\int_D \Omega = \int_C \Omega'$.

Proof. It is only necessary to prove the theorem for the case that C is a simplex. In this case the formula is, however, easily verified.

It might be interesting to consider some particular cases of this theorem. In fact we have

$$(1) \quad \Omega' = \left(\frac{\delta Q}{\delta x} - \frac{\delta P}{\delta y} \right) dxdy, \qquad \text{if } \Omega = P(x,y)dx + Q(x,y)\,dy\ ;$$

$$(2) \quad \Omega' = \left(\frac{\delta Q}{\delta x} - \frac{\delta P}{\delta y} \right) dxdy + \left(\frac{\delta R}{\delta y} - \frac{\delta Q}{\delta z} \right) dydz + \left(\frac{\delta P}{\delta z} - \frac{\delta R}{\delta x} \right) dzdx\ ;$$
$$\text{if } \Omega = P(x,y,z)dx + Q(x,y,z)dy + R(x,y,z)dz\ ;$$

$$(3) \quad \Omega' = \left(\frac{\delta P}{\delta x} + \frac{\delta Q}{\delta y} + \frac{\delta R}{\delta z} \right) dxdydz, \text{ if } \Omega = P(x,y,z)dydz + Q(x,y,z)dzdx + R(x,y,z)dxdy.$$

In this case we find, as special cases of the last theorem, the well known formulæ of Green and Stokes.

6. CONCLUDING REMARKS

There are different ways of forming a set of axioms for the Grassmann and differential rings. The one adopted here seems to be the best suited to our purpose. A set of axioms which fulfils the requirement of rigor and elegance in the spirit of modern abstract algebra has been given by C. Chavalley in an unpublished manuscript.

The multiple integral of an exterior differential form over a simplicial domain can only be defined in a rigorous way, when use is made of the notions of combinatorial topology. As no attempt has been made to go into formal definitions, our treatment is bound to be unsatisfactory so far as rigor is concerned. But we hope that the main ideas have been made clear, so that the reader can supply the details needed to his own judgement of rigor.

In utilising this notion of multiple integral we must not fail to mention that the orientation of the domain plays an important part in the definition. It follows that such a multiple integral is by definition a signed quantity. In this respect it differs essentially from the integral which results from the measure theory, for instance, a Lebesgue integral. It is therefore important to remark that when the absolute value of the integral is under consideration, we must divide the integral into its positive and negative parts and add up their absolute values, not their signed values.

The theory of Grassmann and differential rings has played an important role in the theory of differential systems, in differential geometry, and in topology. It was extensively and successfully employed by E. Cartan (1922 & 1937). Recently it has been proved to be useful in integral geometry, which emerged in mathematics first as an extension of the theory of geometrical probabilities, but is now reaching the stage of an independent branch. We give in the bibliography a list of works in which interesting applications of the theory can be found.

Before concluding we wish to mention one interesting fact. The Grassmann ring of dimension two over the field of real numbers consists of elements of the form $\alpha+\beta x$, such that $x^2=0$. Such an element is called a dual number by E. Study. The dual numbers can be employed to establish certain analogies between line geometry in space and plane euclidean geometry and to study the so-called "equilong transformations" in the plane.

REFERENCES

1. Blaschke, W. (1937) Integral geometrie, I, II. Berlin.
2. Cartan, E. (1922) Lecons sur les invariants integraux. Paris .
3. Cartan, E. (1937) La theorie des groupes finis et continus et la geometrie differentielle traitees par la methode du repere mobile. Paris.
4. Chevalley, C. Unpublished manuscript.
5. de Rham. (1931) Uber mehrfache integrale, Abhandlungen aus dem mathematischen Seminar der Hansischen Universitat,–Bd. 12, 313-339.
6. Forder, H. G. (1941) Calculus of extension. Cambridge.
7. Hodge, W. V. D. (1941) The theory and applications of harmonic integrals. Cambridge.
8. Kahler, E. (1924) Einfuhrung in die theoric der systeme von differentialgleichungen. Berlin.
9. Klein, F. (1926) Vorlesungen über höhere geometrie. Dritte Auflage. Berlin.
10. Thomas, T. M. (1937) Differential systems. New York.
11. Van der Waerden, B. L. (1937) Moderne Algebra I, Zweite Auflage. Berlin.
12. Weil, Andre (1937) Les formes differentielles exterieures, seminaire sur les travaux scientifiques de M. Cartan. Paris.

Paper received : 20 October, 1943.

GENERALISATION OF MARKOFF'S THEOREM AND TESTS OF LINEAR HYPOTHESES

BY C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta

INTRODUCTION

The general theory of linear estimation, without involving the assumpions of normality or independence of variates was first given by Markoff in his book *Calculus of Probability* published in Russian. Sheppard (1912 and 1914) working independently published some results which are approximately on the same lines as that of Markoff, but of a less general character. A distinct advance was made by Aitken (1935) who removed the unnecessary limitations in both Sheppard's and Markoff's results. Recently David and Neyman (1938) have published an article giving a slight extension of Markoff's theorem.

A significant step in generalising the theory of linear estimation is due to Raj Chandra Bose (1943), who, for the first time, introduced the concept of *nonestimable parametric functions*. He distinguishes two types of linear functions of stochastic variates; the estimating functions and the error functions. A linear function $BY' = (b_1y_1 + b_2y_2 + \ldots + b_ny_n)$ of the stochastic variates is said to belong to error if $E(BY') = 0$. The totality of the independent vectors such as B constitute a vector space which is called the error space. The vector space orthogonal to this error space is called the estimation space, and *the best unbiassed estimate of any estimable parametric function* comes out as the scalar product of the vector $Y = (y_1, y_2, \ldots, y_n)$ of the stochastic variates and a vector C of the estimation space. The present paper is the result of ideas suggested by Bose's results (1943) and his post-graduate lectures in the Calcutta University in 1943-44 on the general theory of linear estimation and the fundamental structure of the analysis of variance.

The object of the present paper is firstly to take up the most general problem in the theory of linear estimation and get suitable generalisations of the previous results and secondly to derive tests of significance connected with linear hypotheses.

2. THE GENERAL PROBLEM OF LINEAR ESTIMATION.

Let $Y = (y_1\ y_2 \ldots y_n)$ be the vector of n stochastic variates $y_1, y_2, \ldots, y_n$ and $\Theta = (\theta_1\ \theta_2 \ldots \theta_n)$ be the vector of expectations of $y_1, y_2, \ldots, y_n$. It is given that $\Theta = TA'$ where $T = (\tau_1\ \tau_2 \ldots \tau_m)$ is a row matrix of m unknown parameters and A is a known matrix with n rows and m columns. The transpose of a matrix is denoted by a dash. *Here we do not assume any equality or inequality relations between n and m or make any restrictions on the rank of A.* The rank of A is, evidently, less than or equal to the smaller of m and n. Assuming that there are no functional relationships among the y's we can get a positive definite matrix Λ of rank n as $\Lambda = E(Y - \Theta)'(Y - \Theta)$ where E stands for the mathematical expectation in which case Λ is referred to as the dispersion matrix of the stochastic variates of Y. We assume that Λ is known apart from a constant multiplier and that its elements are finite.

The set of equations $E(Y) = TA'$ are known as observational equations and a linear function such as LT' where L is a given row matrix is called a parametric function.

The problem of linear estimation is to find a linear function BY' of the stochastic variates such that

$$\text{(a)} \qquad E(BY') = LT' \quad \text{independently of } T \tag{2.1}$$

and

$$\text{(b)} \qquad V(BY') \quad \text{is minimum} \tag{2.2}$$

where V stands for the variance. If there exists a vector B such that (a) is satisfied then LT' is said to be *estimable* and BY' is called an *unbiassed estimate* of LT'. If further, (b) is satisfied then BY' is called the best *unbiassed estimate of LT'*.

3. THE GENERAL SOLUTION OF THE PROBLEM OF LINEAR ESTIMATION.

The definition of unbiassedness involved in (2.1) gives that if,

$$LT' \equiv E(BY') \equiv BE(Y') \equiv BAT' \tag{3.1}$$

then, $L = BA$. Also if B is such that $L = BA$ then $LT' = BAT' = E(BY')$. This gives the result : *the necessary and sufficient condition that LT' is estimable is that there exists a vector B such that $L = BA$.*

From the set of B's satisfying the condition $L = BA$ we have to choose B such that $V(BY')$ is the least.

$$V(BY') = E[B(Y' - \Theta')(Y - \Theta)B'] = B\Lambda B' \qquad .. \quad (3.2)$$

Introducing the vector $2C = 2(c_1 c_2 \ldots . c_m)$ of Lagrangian multipliers we have to minimise $B\Lambda B' - 2C(A'B' - L')$ with respect to the elements of B and C. This leads to

$$B\Lambda = CA' \quad \text{and} \quad L = BA \qquad .. \quad (3.21)$$

Defining Λ^{-1} as the inverse of Λ we get

$$B\Lambda\Lambda^{-1} = CA'\Lambda^{-1} \quad \text{or} \quad B = CA'\Lambda^{-1} \qquad .. \quad (3.22)$$

Using the relation $L = BA$ we get, $L = BA = CA'\Lambda^{-1}A$ which is another form of the necessary and sufficient condition. If C *is found to satisfy $L = CA'\Lambda^{-1}A$ then $B = CA'\Lambda^{-1}$ and the best unbaissed estimate of LT' is given by $BY' = CA'\Lambda^{-1}Y'$.*

We prove here some uniqueness results. If C_1 and C_2 are two vectors satisfying the relation $L = CA'\Lambda^{-1}A$, then $0 = (C_1 - C_2)A'\Lambda^{-1}A$. If B_1 and B_2 are the two vectors derived from (3.22) corresponding to C_1 and C_2 then $(B_1 - B_2) = (C_1 - C_2)A'\Lambda^{-1}$. Since

$$\begin{aligned} V(B_1 - B_2\, Y') &= (C_1 - C_2)A'\Lambda^{-1}\Lambda\Lambda^{-1}A(C_1 - C_2)' \\ &= (C_1 - C_2)A'\Lambda^{-1}A(C_1 - C_2)' = 0 \quad (C_1 - C_2)' = 0 \end{aligned} \qquad .. \quad (3.3)$$

we get that $B_1 \equiv B_2$ for the variance of a linear function of stochastic variates cannot be zero without the coefficients being zero. This leads to the result *that the vector B or the best unbiassed estimate BY' is unique for all C's satisfying the relation $L = CA'\Lambda^{-1}A$.*

If B is the vector derived in (3.22) and D is any other vector such that $L = DA$ and is also a best unbiassed estimate of LT' then

$$\begin{aligned} V(DY') = D\Lambda D' &= B\Lambda B' + (D - B)\Lambda B' + B\Lambda(D' - B') + (B - D)\Lambda(B' - D') \\ &= B\Lambda B' + (B - D)\Lambda(B' - D') \end{aligned} \qquad .. \quad (3.4)$$

for

$$B\Lambda(D' - B') = (D - B)\Lambda B' = (D - B)\Lambda\Lambda^{-1}AC' = (D - B)AC' = 0 \qquad .. \quad (3.41)$$

$(B - D)\Lambda(B' - D')$ being positive definite cannot be negative. Hence the least variance is given when $B - D = 0$, which leads to the result *that the calculations involved in* (3.22) *give a unique vector B such that $V(BY')$ is the least.*

4. EXTENSION OF MARKOFF'S PRINCIPLE

From Y, we construct the vector $Q=(Q_1\ Q_2 \ldots . Q_m)$ composed of *m* linear functions of y's defined by $Q=Y\Lambda^{-1}A$. If LT' is estimable then we have $L=CA'\Lambda^{-1}A$ which gives that $LT'=CA'\Lambda^{-1}AT'=CH'$ where,

$$H=E(Q)=E(Y)\Lambda^{-1}A=TA'\Lambda^{-1}A \qquad .. \quad (4.1)$$

Since the best unbiassed estimate is given by $BY'=CA'\Lambda^{-1}Y'=CQ'$ we can restate the result of section 2 as, *if there exists a vector C such that $LT'=CH'$, then LT' is estimable and the best unbiassed estimate of LT' is given by CQ' where H and Q are as defined in* (4.1) *and it is unique for all C's satisfying relation $LT'=CH'$.*

Let $\bar{T}=(t_1\ t_2 \ldots . t_m)$ be a solution of the equations $Q=H$. Since $LT'\equiv CH'$ we get $L\bar{T}'=C\bar{H}'=CQ'$ where $\bar{H}$ denotes the matrix obtained from H by replacing T by $\bar{T}$. So we get the following rule of *estimating* any *estimable* parametric function. *We get a solution of the equations $Q=H$ solved for τ's, and substitute the t's for τ's in the given parametric function. This gives the best unbiassed estimate of LT' and is unique for all solutions of τ's.* The proof of the latter part of the statement follows from the fact that whatever may be the solution, so long as H gives Q we get $L\bar{T}'=CQ'$ where $LT'\equiv CH'$.

We shall now study the effect of substituting a solution of $Q=H$ in a non-estimable parametric function NT'. If $\bar{T}_1$, and $\bar{T}_2$ are two solutions of $Q=H$, such that $\bar{T}_1\neq\bar{T}_2$, then $E(N\bar{T}'_1)\neq E(N\bar{T}'_2)$. For if $E(N\bar{T}'_1)=E(N\bar{T}'_2)$ then

$$E[N(\bar{T}'_1-\bar{T}'_2)]=0 \quad \text{or} \quad N(\bar{T}'_1-\bar{T}'_2)=0 \qquad .. \quad (4.2)$$

for $(\bar{T}_1-\bar{T}_2)$ being a non-null solution of the homogeneous equations $H=0$, is independent of Q's and hence, of the stochastic variates. The result (4.2) shows that N belongs to the matrix of equations $H=0$ or $N=CA'\Lambda^{-1}A$ which means that NT' is estimable contrary to our supposition. Again, if $\bar{T}$ is a solution of $Q=H$, and if $N\bar{T}'$ is homogeneous in $y_1, y_2 \ldots, y_n$ then $E(N\bar{T}')=NT'$ when and only when NT' is estimable for otherwise it means that there exists a linear homogeneous function of Y's such that its expectation is NT'. Also the result of substitution of $\bar{T}$ in an estimable parametric function NT' leads to a homogeneous expression in y's. So we get *that the result of substitution of any solution* $\bar{T}$ of *$Q=H$ in a parametric function LT' leads to the best unbiassed estimate of LT' if and only if*

$$\begin{array}{ll} \text{(i)} & L\bar{T}' \text{ is homogeneous in y's} \\ \text{(ii)} & E(L\bar{T}')=LT' \end{array} \qquad .. \quad (4.21)$$

A non-estimable parametric function will show forth *either in violation of* (*i*) *or* (*ii*) *or in giving two different expressions for two different solutions of $Q=H$.* We need not test for estimability before applying Markoff's principle of substitution. The above discussion shows that we can add a consistent, and a convenient and sometimes conventionally choosen set of equations in τ's not necessarily linear to $Q=H$ and get a solution for substitution.

The equations $Q=H$ are called normal equations and are readily obtained by equating the partial derivatives of

$$\sum_i \sum_j \lambda^{ij}(y_i-\theta_i)\ (y_j-\theta_j) \qquad .. \quad (4.3)$$

where $\lambda^{ij}(i, j=1, 2, \ldots, n)$ are elements of the matrix Λ^{-1}, with respect to $\tau_1, \tau_2 \ldots, \tau_m$, to zero. *The above equations $Q=H$ are always solvable.* To prove this it is enough to show that if D is a vector such that $DH' \equiv 0$, then $DQ' \equiv 0$. $DH' \equiv 0$ implies that $DA'\Lambda^{-1}A = 0$ and

$$V(DQ') = V(DA'\Lambda^{-1}Y) = DA'\Lambda^{-1}\Lambda\Lambda^{-1}AD' = DA'\Lambda^{-1}AD' = 0D' = 0 \qquad .. \quad (4.4)$$

which shows that $DQ' \equiv 0$ if, $DH' \equiv 0$.

We now prove some results which give the number of estimable and non estimable parametric functions.

(a) If $NQ' \equiv 0$, then the parametric function NT' is not estimable. For, if NT' is estimable then $N = CA'\Lambda^{-1}A$ and $NN' = NA'\Lambda^{-1}AC' = 0$ for $NQ' \equiv 0$ implies that $NH' \equiv 0$ and this $(NN' = 0)$ is impossible unless $N = 0$.

(b) There cannot be any non-vanishing linear function of Q's whose expectation is zero for it can be easily shown that the variance of such a linear function, if it exists, should be zero.

(c) The rank of $A'\Lambda^{-1}A$ is the same as the number of Q's which are linearly independent, and also of the number of linear parametric functions that are estimable. This follows from (a) and (b) given above. Since the rank of $A'\Lambda^{-1}A$ is the same as that of A it follows that the number of estimable parametric functions is equal to 's' the rank of A which is less than or equal to the smaller of m and n. The number of non-estimable parametric functions is therefore, $(m-s)$. If the rank of A is m, then all parametric functions are estimable.

5. INTRINSIC PROPERTIES OF NORMAL EQUATIONS.

The normal equations are $Q = TA'\Lambda^{-1}A$. The dispersion matrix of $Q_1, Q_2, \ldots, Q_m$ is

$$\begin{aligned} E[(Q'-H')(Q-H)] &= E[A'\Lambda^{-1}(Y'-\Theta')(Y-\Theta)\Lambda^{-1}A] \\ &= A'\Lambda^{-1}\Lambda\Lambda^{-1}A = A'\Lambda^{-1}A \qquad .. \quad (5.1) \end{aligned}$$

which shows that variance (Q_i) = coefficient of τ_i in the i-th normal equation and covariance $(Q_i\, Q_j)$ = coefficient of τ_j in the i-th normal equation.

If $L_k\, T'$ is estimated by $C_k\, Q', (k = 1, 2, \ldots. p)$ then

$$\begin{aligned} V(C_iQ') &= C_iA'\Lambda^{-1}AC'_i = L_iC_i' \\ \text{Cov}\,[(C_iQ')\,(C_jQ')] &= C_iA'\Lambda^{-1}AC'_j = L_iC'_j = L_jC'_i \qquad .. \quad (5.2) \end{aligned}$$

Let us eliminate τ_1 from the equations $Q = TA'\Lambda^{-1}A$ by the usual method of sweep-out. We divide both sides of the first equation by the coefficient of τ_1 and subtract it, after multiplying it by the coefficient of τ_1 in the i-th equation from the i-th equation. We now get $(m-1)$ equations which can be represented by $Q_1 = T_1B'$. It is easy to show that these equations also satisfy the properties of normal equations given in (5.1) and (5.2) above. The dispersion matrix of Q_1's is B and if $L_{i1}\, T'_1$ is estimated by $C_{k1}\, Q'_1$ then

$$\begin{aligned} V(C_{i1}\, Q'_1) &= L_{i1}\, C'_{i1} \\ \text{Cov}\,[(C_{i1}\, Q'_1)\,(C_{j1}\, Q'_1)] &= L_{i1}\, C'_{j1} = L_{j1}\, C'_{i1} \qquad .. \quad (5.3) \end{aligned}$$

These results hold good even if we eliminate more than one τ, the resulting equations satisfy these intrinsic properties of normal equations.

6. THE NATURE OF LINEAR HYPOTHESES.

The hypothesis involved in the theory of linear estimation is the assignment of the values of a single parametric function or a set of parametric functions. The necessary statistics and their distributions when the stochastic variates form a multivariate normal system are discussed below. The problem of distributions for other types of populations will be discussed in a future communication.

If LT′ is an estimable parametric function with a specified value ξ and its best unbiassed estimate is CQ′ then to test the hypothesis $LT' = \xi$ we construct the statistic

$$v = (CQ' - \xi) / \sqrt{CL'} \qquad .. \quad (6.1)$$

which becomes a normal deviate with unit variance. We can extend this result to test the composite hypothesis $L_i T' = \xi_i (i = 1, 2, .. k < s)$. Let $P = (P_1 \; P_2 \ldots . P_k)$ be the vector giving the estimates of $L_1 T' - \xi_1$, $L_2 T' - \xi_2$, ... $L_k T' - \xi_k$ with the dispersion matrix of P's as D. Following Fisher's (1939) technique we take GP′ a linear compound of P's, where $G = (g_1 \; g_2 \ldots . g_k)$ with its variance GDG′ and maximise the statistic

$$v^2 = (GP')^2 / GDG' \qquad .. \quad (6.2)$$

If we denote by V the maximum value of v^2 we get V as the root of the determinental equation $|P'P - VD| = 0$. This gives $V = \Sigma \; \Sigma d^{ij} P_i \; P_j$ where d^{ij} are the elements of matrix D^{-1}, the inverse of D. This may be called the *generalised variance statistic*. This statistic is invariant for any set of relations derived by linear combinations of the relations $L_i \; T' - \xi_i$. For in every case, we will be led to maximise the expression

$$V = (GP')^2 / \bar{G}D\bar{G}' \qquad .. \quad (6.21)$$

where the elements of G are linear combinations of the elements of G with which we started.

Under the normality assumption we get the distribution of P's as

$$\text{Const.} \quad e^{-\frac{1}{2}\psi} . \pi . dP_i \qquad .. \quad (6.3)$$

where

$$\psi = \Sigma\Sigma d^{ij} \; P_i \; P_j \qquad .. \quad (6.31)$$

from which it immediately follows that the distribution of V is

$$\text{Const.} \quad [exp \; (V/2] \; V^{(k-2)/2} \; dV \qquad .. \quad (6.32)$$

which is the distribution of χ^2 with k degrees of freedom and it supplies the distribution of the V-statistic on the null hypothesis.

If the quantities $\phi_i = L_i \; T' - \xi_i \neq 0$ for at least a single i, then there is departure from the null-hypothesis. We define the quantity

$$\phi = \Sigma\Sigma d^{ij} \; \phi_i \; \phi_i \qquad .. \quad (6.4)$$

which is also invariant for any linear combinations of ϕ's. If we transform the estimating functions P's to R's by linear combinations to make them independent and having unit variances the distribution (6.3) transforms to

$$\text{Const.} \quad e - \tfrac{1}{2} \psi \; dR_1 \; dR_2 \ldots . dR_k \qquad .. \quad (6.41)$$

where

$$\psi = (R_1 - \rho_1)^2 + (R_2 - \rho_2)^2 + \ldots . + (R_k - \rho_k)^2 \qquad .. \quad (6.42)$$

and ρ's are linear combinations of ϕ's. V and ϕ transform to $R = \Sigma R^2$ and $\rho = \Sigma \rho_i^2$. By an argument similar to Bose (1935) we get the distribution of R or V as

$$\text{Const.} \quad [exp(V/2)] \; V^{(k-2)/2} I_{k-2}(\sqrt{V\phi}) dV \qquad .. \quad (6.43)$$

where I is the Bessel function of the second kind. This is the distribution of the generalised variance statistic V on the non null hypothesis and is dependent on k and ϕ only.

We consider here a class of linear hypotheses which are of special interest in tests of significance connected with biometrics and field experimentation. These are no doubt, special cases of the general composite hypothesis discussed above, but are amenable to a logical deduction of the necessary statistic. The hypothesis consists in specifying the equality of a number of parametric functions such as

$$L_1 T' = L_2 T' = \ldots . = L_k T' \qquad \ldots \quad (6.5)$$

If $P_1, P_2, \ldots, P_k$ are the estimates of these linear functions then we take the linear compound

$$g_1 P_1 + g_2 P_2 + \ldots . + g_k P_k \qquad \ldots \quad (6.51)$$

such that $g_1 + g_2 + \ldots . + g_k = 0$ and maximise the expression

$$v^2 = (GP')^2 / GDG' \qquad \ldots \quad (6.52)$$

This comes to testing for a suitable contrast of the parametric functions $L_i T'$. The maximisation leads to the determinental equation

$$|(P_i - \bar{P})(P_j - \bar{P}) - VD| = 0 \qquad \ldots \quad (6.53)$$

where

$$\bar{P} = \sum_i P_i (\sum_j d^{ij}) / \sum_i \sum_j d^{ij} \qquad \ldots \quad (6.54)$$

which gives

$$V = \Sigma\Sigma d^{ij} (P_i - \bar{P})(P_j - \bar{P}) \qquad \ldots \quad (6.55)$$

The distribution of this follows directly from previous considerations for it is designed to test a hypothesis concerning (k−1) independent parametric functions. Hence V follows the distribution

$$\text{Const.} \ [exp(V/2)] \ V^{(k-3)/2} dV \qquad \ldots \quad (6.56)$$

on the null hypothesis and

$$\text{Const.} \ [exp(V/2)] V^{(k-3)/2} \ I_{(k-3)/2} (\sqrt{V}\phi) dV \qquad \ldots \quad (6.57)$$

on the non-null hypothesis, ϕ being defined as

$$\phi^2 = \Sigma \ \Sigma d^{ij} (\theta_i - \bar{\theta})(\theta_j - \bar{\theta}) \qquad \ldots \quad (6{\cdot}58)$$

where θ_i stands for $L_i T'$ and $\bar{\theta}$ is the same functions of θ_i's as $\bar{P}$ is of P_i's.

In the case of the multinomial distribution with the proportions $\pi_1, \pi_2, \ldots \pi_k$ in the k classes the linear observational equations corresponding to a sample of $n_1, n_2, \ldots n_k$ $(\Sigma n_i = N)$ in the k classes are $E(n_i) = N\pi_i (i = 1, 2, \ldots k)$. We want to test the hypothesis whether the sample is a reasonable one from the population with the assigned proportions $\pi_1, \pi_2, \ldots . \pi_k$. The observational equations $E(n_i) = N\pi_i (i = 1, 2, \ldots k-1)$ which may be taken to be independent will themselves be estimating equations. Using the properties

$$V(n_i) = \pi_i (1 - \pi_i) N \qquad \ldots \quad (6.6)$$

$$\text{Cov} \ (n_i \ n_j) = -\pi_i \pi_j N \qquad \ldots \quad (6.61)$$

we try to maximum the statistic

$$v = \sum_{i=1}^{k-1} g_i (n_i - N\pi_i) / [\Sigma g^2_i \ \pi_i (1 - \pi_i) N - \Sigma\Sigma g_i \ g_j \ \pi_i \ \pi_j \ N]^{\frac{1}{2}} \qquad \ldots \quad (6.62)$$

The maximum value comes out as

$$V = \sum_{i=1}^{k} (n_i - N\pi_i)^2 / N\pi_i \qquad \ldots \quad (6\ 63)$$

which is the same as the χ^2-statistic of Karl Pearson.

The discussion in the above paragraphs form the basis of the perimeter test derived by the author and discussed in another paper (1944). There are t, p-variate populations with means given by the matrix $((m_{ij}))$ $i=1, 2, \ldots . p$ referring to the variates of a population $j=1, 2, \ldots . t$ referring to the populations, the hypothesis to be tested is that

$$m_{i1}=m_{i2} \ldots . = m_{it}, \; (i=1, 2, \ldots . p) \qquad .. \quad (6.7)$$

There are $p(t-1)$ independent parametric functions to be tested. Making use of previous results we get the generalised variance statistic as

$$V = \sum_{j=1}^{t} \sum_{r=1}^{p} \sum_{s=1}^{p} {}_j\alpha^{rs}(\bar{x}_{rj}-\bar{x}_r)\;(\bar{x}_{sj}-\bar{x}_s)n_j \qquad .. \quad (6.71)$$

where ${}_j\alpha^{rs}$ are the elements reciprocal to the variance and covariance matrix for the j-th p-variate population from which a sample of size n_j is drawn and $\bar{x}_{rj}$ is the sample mean for the r-th character in the j-th population and $\bar{x}_r$ is defined by

$$(-1)^r\,\bar{x}_r \begin{vmatrix} c_{11} & c_{12} & \ldots . & c_{1p} \\ c_{21} & c_{22} & \ldots . & c_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ c_{p1} & c_{p2} & \ldots . & c_{pp} \end{vmatrix} = \begin{vmatrix} c_{11} & \ldots & c_{1r-1} & d_1 & c_{1\,r+1} & \ldots & c_{1p} \\ c_{21} & \ldots & c_{2r-1} & d_2 & c_{2\,r+1} & \ldots & c_{2p} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ c_{p1} & \ldots & c_{pr-1} & d_p & c_{p\,r+1} & \ldots & c_{pp} \end{vmatrix} \qquad .. \quad (6.72)$$

where $$d_s = \sum_{j=1}^{t} \sum_{l=1}^{p} {}_j\alpha^{ls}\,\bar{x}_{lj}\; n_j \quad \text{and} \quad c_{lm} = \sum_{j=1}^{t} n_j\; {}_j\alpha^{lm} \qquad .. \quad (6.73)$$

The distribution of this statistic on the null and non-null hypotheses are obtained directly from (6.32) and (6.43) by putting $k=p(t-1)$.

An important problem in the classification of three p variate populations π_1, π_2 and π_3 is to test whether π_1 is nearer to π_2 or π_3 which are known to be different. If m_{ij} is the mean of the j-th character in the i-th population we have to test the composite hypothesis

$$2m_{1j}=m_{2j}+m_{3j} \quad j=(1, 2, \ldots . p) \qquad .. \quad (6.8)$$

If $\bar{x}_{ij}$ is the sample mean for the j-th character of the i-th population and n_i is the sample size for the i-th population then the necessary statistic is

$$V = \Sigma\; \Sigma d^{ij}(2\bar{x}_{1j}-\bar{x}_{2j}-\bar{x}_{3j})\;(2\bar{x}_{1i}-\bar{x}_{2i}-\bar{x}_{3i}) \qquad .. \quad (6.81)$$

where α^{ij} are the elements of the matrix reciprocal to

$$\left(\left(4\,\frac{\alpha_{ij}^{(1)}}{n_1} + \frac{\alpha_{ij}^{(2)}}{n_2} + \frac{\alpha_{ij}^{(3)}}{n_3}\right)\right) \qquad .. \quad (6.82)$$

where α^k_{ij} is the covariance of the i-th and the j-th characters in the k-th population. The distribution of V on the null and non-null hypotheses are obtained from (6.32) and (6.43) by putting $k=p$.

The necessary statistics when the variances and the covariances are not known are obtained by studentising the above statistics and their distributions in the studentised forms have been derived and are being used in the Statistical Laboratory at Calcutta for various problems. A detailed discussion will be attempted in another paper.

REFERENCES

1. AITKEN, A. C. (1935) On the least square and linear combination of observations. *Proc. Roy. Soc Edin.* **55**, 42-48.

2. BOSE, R.C. (1944) The fundamental theorem of linear estimation. *Proc. Ind. Sc. Congress.* 4(3).

3. DAVID, F. N. and NEYMAN, J. (1939) Extension of Markoff's theorem on least squares. *Biom. Research Memoirs.* **2**, 103-116.

4. FISHER, R. A. (1939) The sampling distribution of some statistics obtained from nonlinear regression. *Annals of Eugenics.* **9**, 238-249.

5. MARKOFF, A. A. (1904) Calculus of Probability. Russian Edition.

6. RAO, C. R. (1944) The generalised variance of populations. *Proc. Ind. Sc. Congress.* 4 (3)

7. SHEPPARD, W. F. (1912) Reduction of errors by negligible differences. *Proc. Fifth Int. Congr. Math (Cambridge)* **2**, 348-84.

8. SHEPPARD, W. F. (1914 a) Fitting of polynomials by method of least squares. *Proc. Lond. Math. Soc.* **13** (3), 97-108.

9. SHEPPARD, W. F. (1914 b) Graduation by a reduction of mean square of error. *Jour. Inst. Actuaries.* **48**, 171-185, 390-412. and **49**, 148-157.

Paper received : 20 January, 1944

MARKOFF'S THEOREM WITH LINEAR RESTRICTIONS ON PARAMETERS

By C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta

INTRODUCTION

In an earlier paper by the author (1945) the problem of linear estimation has been dealt with under very general conditions. The problem is, given the vector $Y = (y_1 \; y_2 \; \ldots \; y_n)$ of n stochastic variates such that $E(Y) = TA'$ where $T = (\tau_1 \; \tau_2 \; \ldots \; \tau_m)$ is a row matrix of m unknown parameters and A is a known matrix with n rows and m columns and also the dispersion matrix Λ of the stochastic variates, to find a linear function BY′ such that

$$\text{(a)} \quad E(BY') = LT' \quad \text{independently of T} \qquad (1{\cdot}1)$$

and

$$\text{(b)} \quad V(BY') \quad \text{is minimum} \qquad (1{\cdot}11)$$

where LT′ is a given parametric function.

The object of the present paper is to extend the above problem to cases where the parameters $\tau_1, \tau_2, \ldots, \tau_m$ are subject to k linear restrictions $G = TR'$ where G and R are known matrices the former being a row matrix with k elements and the latter being a matrix with k rows and m columns. The problem is to find a linear function $b_0 + BY'$ (if one exists, not necessarily homogeneous in y's) such that

$$\text{(a)} \quad E(b_0 + BY') = LT' \quad \text{subject to } G = TR' \qquad (1{\cdot}2)$$

$$\text{(b)} \quad V(b_0 + BY') \quad \text{is minimum} \qquad (1{\cdot}21)$$

The discussion on tests of linear hypotheses given by Rao (1945) holds good in these situations also.

2. SOLUTION TO THE GENERAL PROBLEM

If b_0 and B exist such that (1.2) is satisfied then

$$LT' = E(b_0 + BY') = b_0 + BAT' \qquad \cdot\cdot \quad (2\cdot 1)$$

is true whenever $G' = RT'$ which gives that for a suitable vector D we get

$$b_0 + (BA - L)\,T' \equiv D(G' - RT') \qquad \cdot\cdot \quad (2\cdot 11)$$

$$\text{or} \quad BA + DR = L \quad \text{and} \quad b_0 = DG' \qquad \cdot\cdot \quad (2\cdot 12)$$

Also if b_0 and B satisfy (2·12) then $E(b_0 + BY') = LT'$ which gives that *the necessary and sufficient condition that LT' is estimable is that there exist b_0 and B such that for a suitable D the relations $BA + DR = L$ and $b_0 = DG'$ are identically true.*

From the set of b_0, B and D satisfying (2.12) we have to select those for which $V(BY') = B\Lambda B'$ is the least. Introducing Lagrangian multipliers $2d$ and $2C = 2(c_1, c_2, \ldots, c_m)$ we have to minimise $B\Lambda B' - 2C(A'B' + R'D' - L') + 2d(b_0 - DG')$ with respect to d, b_0 and the elements of B, D and C. This leads to

$$B\Lambda - CA' = 0, \qquad BA + DR = L \qquad \cdot\cdot \quad (2\cdot 2)$$

$$CR' = dG, \quad d = 0, \quad b_0 = DG' \qquad \cdot\cdot \quad (2\cdot 21)$$

Defining Λ^{-1} as the inverse of Λ we get

$$B = CA'\Lambda^{-1}, \qquad CA'\Lambda^{-1}A + DR = L \qquad \cdot\cdot \quad (2\cdot 22)$$

$$CR' = 0, \quad b_0 = DG' \qquad \cdot\cdot \quad (2\cdot 23)$$

So we get *that b_0 and* B *satisfying* (1.2) *and* (1.21) *are given by* $B = CA'\Lambda^{-1}$ *and* $b_0 = DG'$ *where* C *and* D *satisfy the equations*

$$CA'\Lambda^{-1}A + DR = L, \; CR' = 0 \qquad \cdot\cdot \quad (2\cdot 24)$$

As shown earlier (Rao : 1945) we can easily prove that b_0 and B are unique for all C and D satisfying (2.24) and the variance associated with $b_0 + BY'$ derived from above is the least.

3. THE ANALOGUE OF MARKOFF'S RULE

It has been shown above that the best estimate of LT' is given by $DG' + CA'\Lambda^{-1}Y'$ where C and D satisfy (2.24). If we construct the quantities $Q = (q_1, q_2, \ldots, q_m)$ by the transformation $Q = Y\Lambda^{-1}A$ with $E(Q') = A'\Lambda^{-1}AT' = H'$ we can restate the result of section (2) as, *if there exist* C *and* D *such that* $LT' = CH' + DRT'$ *and* $CR' = 0$ *then the best unbiassed estimate of* LT' *is* $DG' + CQ'$ *and this is unique for all* C *and* D *satisfying the above relation.*

We may translate the above result into the following rule. Let us construct the linear equations

$$Q = H + SR \quad \text{and} \quad CR' = 0 \qquad \cdot\cdot \quad (3\cdot 2)$$

in $\tau_1, \tau_2, \ldots, \tau_m$ and k pseudo-variates $\sigma_1, \sigma_2, \ldots, \sigma_k$ of the vector $S = (\sigma_1, \sigma_2, \ldots, \sigma_k)$ and get $\bar{T} = (t_1, t_2, \ldots, t_m)$ and $\bar{S} = (s_1, s_2, \ldots, s_k)$ as solutions. Because of the identity

$$LT' = DRT' + CH' \quad \text{where } CR' = 0 \qquad \cdot\cdot \quad (3\cdot 21)$$

$$L\bar{T}' = DR\bar{T}' + CH' = DG' + C\,(Q' - R'\bar{S}) = DG' + CQ' \qquad \cdot\cdot \quad (3\cdot 22)$$

we get the result *that the substitution of any solution of* (3.2) *in an estimable parametric function* LT' *leads to its best estimate.*

If NT′ is any given parametric function then E(NT′) = NT′ when and only when NT′ is estimable. This may be used as the criterion of estimability. A non-estimable parametric function will reveal itself either in violation of this criterion or in giving different expressions for different solutions.

The equations (3.2) may be called normal equations appropriate for this case and are readily obtained by minimising the expression.

$$\Sigma\,\Sigma\,\lambda^{ij}\,(y_i-\theta_i)\,(y_j-\theta_j) \quad .. \quad (3\cdot 4)$$

where λ^{ij} $(i, j = 1, 2, \ \ldots, n)$ are the elements of Λ^{-1}, with respect to $\tau_1, \tau_2, \ldots \tau_m$ subject to the restrictions G = TR′ thus leading to the usual rule of least squares.

The equations (3.2) are always solvable. For this it is enough to show that if there exist U and W such that

$$U(H' + R'S') + WRT' \equiv 0 \quad .. \quad (3\cdot 5)$$

then UQ′+DG′≡0 which is true for we can show that both the expectation and the variance of UQ′+DG′ vanish if (3.5) holds.

The above discussion shows that we can add a consistent and a convenient and sometimes a conventionally chosen set of equations to (3.2) and get solutions for substitution.

From the necessary and sufficient condition for estimability (2.12) we derive that the number of estimable parametric functions is equal to the rank of the matrix obtained by adjoining R to A.

4. VARIANCES AND COVARIANCES OF ESTIMATES

The best estimate of LT′ is given by DG′ +CQ′ where C and D satisfy (2·24). It follows that

$$V(DG' + CQ') = CA'\Lambda^{-1}AC' = (L - DR)C' = LC' \quad .. \quad (4\cdot 1)$$

If LT′ and MT′ are estimated by DG′+CQ′ and ΔG′+χQ′, then

$$\text{cov.}\,(DG' + CQ')\,(\Delta G' + \chi Q') = \text{cov.}\,(CQ')\,(\chi Q') = L\chi' = MC' \quad .. \quad (4\cdot 11)$$

which are analogous to those derived by Rao in (1945).

We can also show that the intrinsic properties of normal equations, discussed in the previous paper, hold good in this case also. These results, in particular cases, have been already discussed by the author elsewhere (Rao: 1943).

5. TESTS OF LINEAR HYPOTHESES

The nature of linear hypothesis is the assignment of the value of one or more linear parametric functions. The case of a single parametric function does not present any difficulty. If the y's form a multivariate normal system, we construct the normal variate with zero mean and unit variance by taking the ratio of the deviation of the estimate from the given value to its standard deviation and test for its significance. Thus if LT′ is an estimable parametric function with a specified value ξ and its best unbiassed estimate is DG′+CQ′ then to test the hypothesis LT′ = ξ we construct the statistic

$$\tau = (DG'+CQ'-\xi)/\sqrt{CL'} \quad .. \quad (5\cdot 0)$$

which can be referred to a probability integral table of normal deviates.

We can extend this result to test the composite hypothesis that a number of parametric functions $L_1T', L_2T', \ldots, L_pT'$ have assigned values $\xi_1, \xi_2, \ldots, \xi_p$. If E and ξ be the vectors defined by $(L_1T' \; L_2T' \ldots L_pT')$ and $(\xi_1 \xi_2 \ldots \xi_p)$ then for the hypothesis to be consistent it follows that if there exists two vectors $M=(m_1 \; m_2 \ldots m_p)$ and $N = (n_1 \; n_2 \ldots n_k)$ such that $ME' \equiv NRT'$ or 0, then $M\xi' = NG'$ or 0 respectively. In this case we can replace the composite hypothesis by finding an independent set of parametric functions $M_1T', M_2T' \ldots, M_mT'$ and corresponding values $\eta_1, \eta_2, \ldots, \eta_m$ derived by linear combinations from $L_1T', L_2T', \ldots, L_pT'$ such that no linear combination of the new set can be derived from the parametric functions TR' occuring in the linear restrictions $G=TR'$. If the best estimates of $M_1T'-\eta_1, M_2T'-\eta_2, \ldots, M_mT'-\eta_m$ are given by the elements of the vector $P=(P_1 P_2 \ldots P_m)$ with their dispersion matrix given by D, then following the argument, as in the previous paper, we get the statistic V, suitable for testing the composite hypothesis as the root of the determinental equation

$$|P'P - VD| = 0 \qquad \ldots \quad (5{\cdot}1)$$

The statstic V is distributed as χ^2 with m degrees of freedom on the null hypothesis and as the Bessel function variate defined in the previous paper, on the non-null hypothesis.

Since V is invariant when the matrices $P'P$ and D are pre or post multiplied by non singular square matrices, it follows that V is the same for any set of independent functions $M_1T', M_2T', \ldots, M_mT'$ chosen above.

The necessary statistics when the variances and covariances are not known are obtained by studentising the above statistics with the help of suitable quadratic estimates of the variances and covariances. The general problem of quadratic estimation is being considered for this purpose and it will be shown elsewhere, that many problems concerning linear hypothesis can be answered with the help of two distributions viz the t and F distributions.

REFERENCES

1. RAO, C. R. (1943) Thesis submitted for the degree of M.A. in the Calcutta University.

2. RAO, C. R. (1945) Generalisation of Markoff's theorem and tests of linear hypotheses. *Sankhya*, 7(1), 9-16.

Paper received : 8 May, 1945.

ON THE ACCURACY OF THE DIFFERENT APPROXIMATIONS TO THE L_1-DISTRIBUTION

By Miss ALEYAMA GEORGE
University of Travancore

INTRODUCTION

To test whether a set of k samples can be regarded as coming from p-variate normal populations with the same set of variance and covariance parameters, Wilks (1932) proposed the criterion λ_1 derived by the likleihood method of Neyman and Pearson (1930 and 1931) where λ_1 is given by

$$\lambda_1 = \prod_{\beta=1}^{k} \left| V_{ij} \right|_{\beta}^{\frac{1}{2}n_\beta} \Big/ \left| V_{ij} \right|_{0}^{N/2} \qquad .. \quad (1\cdot1)$$

where $|V_{ij}|_\beta$ is the determinant of the dispersion matrix for the βth sample, n_β the size of the β^{th} sample and $|V_{ij}|_0$ the determinant of the dispersion matrix of the pooled variances and covariances for the k samples. The sampling distribution of λ_1 has not however been determined but various approximations have been proposed. Since the distribution of λ_1 is highly skew as shown by Neyman and Pearson (1930 and 1931) we shall use the criterion,

$$L_1 = \lambda_1^{2/N} = \prod_{\beta=1}^{k} \left| V_{ij} \right|_{\beta}^{n_\beta/N} \Big/ \left| V_{ij} \right|_{0} \qquad .. \quad (1\cdot2)$$

2. DISTRIBUTION OF L_1

In this paper the exact distribution of L_1 has been worked out for the following cases.

A. Equal Samples ($n_k = n$).

(i) $k = 2$, $p = 2$. (ii) $k = 2$, $p = 3$. (iii) $k = 2$, $p = 4$. (iv) $k = 3$, $p = 2$.
(v) $k = 3$, $p = 3$.

B. Unequal samples.

(i) $k = 2$, $p = 3$, $n_1 = n$, $n_2 = 2n$, $N = 3n$.
(ii) $k = 3$, $p = 2$, $n_1 = n$, $n_2 = 2n$, $n_3 = 3n$, $N = 6n$.
(iii) $k = 3$, $p = 2$, $n_1 = n_2 = n$, $n_3 = 2n$, $N = 4n$.

The t-th moment of L_1 is given by

$$\Phi(t) = \prod_{1}^{k} \left[(N/n_\beta)^{p n_\beta t/N} \prod_{i=1}^{p} \Gamma\{n_\beta t/N + (n_\beta - i)/2\} / \Gamma\{(n_\beta - i)/2\} \right]$$
$$\times \prod_{i=1}^{p} \left[\Gamma\{(N-k+1-i)/2\} / \Gamma\{t + (N-k+1-i)/2\} \right] \qquad .. \quad (2\cdot1)$$

Thus by using Mellin's transformation, the distribution of L_1 is given by

$$P(L_1) = \frac{\prod_{i=1}^{p} \Gamma\{(N-k+1-i)/2\}}{\prod_{i=1}^{p}\prod_{\beta=1}^{k} \Gamma\{(n_\beta - i)/2\}} \; \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} L_1^{-t-1} \; \frac{\prod_{\beta=1}^{k} (N/n_\beta)^{pn_\beta t/N} \prod_{i=1}^{p} \Gamma\{n_\beta t/N + (n_\beta - i)/2\}}{\prod_{i=1}^{p} \Gamma\{t+(N-k+1-i)/2\}} \quad \ldots \quad (2{\cdot}2)$$

The exact distribution has been worked out in the cases mentioned above by evaluating the residues at the poles of the integrand when the contour is taken as a semicircle on the half plane given by $x > 0$ and the imaginery axis as diameter.

3. GRAPHS OF P (L_1)

From the exact distribution of L_1, it is evident that the expression for (2.2) becomes very complicated as the values of k or p increase. Further, the functional nature of the solution also cannot be studied with ease for the general case.

The exact distributions serve to give an insight into the nature of the graph of the distribution curve. In all the curves, $P(L_1)$ splits into two factors, one of which is independent of the size of the sample. Thus $P(L_1) = f_1(n, L_1) . f_2(L_1)$.

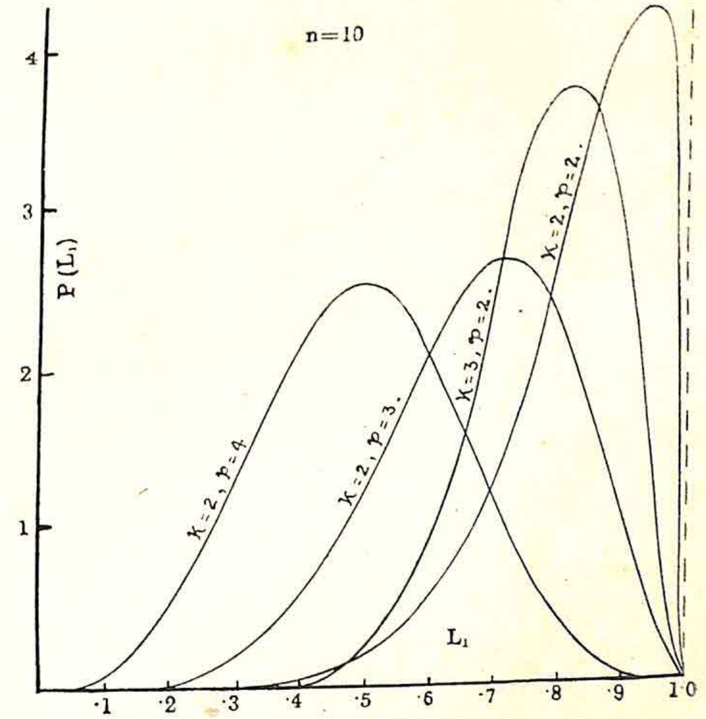


Fig. 1

The curves have been traced for $n = 10$, 20 and 30 and from the graphs it will be observed that, (1) as the value of k or p increases, the curves tend to loose their highly skew nature and more or less tend to normality and (2) there is also an indication that the curves may tend to get centred at $L_1 = 1$ as n increases. It may be pointed out that in practical

8-389
7.9.70

problems, the number of variables may not be large though the number of samples and their sizes may be large. When n is large it has already been pointed out that $\log_e L$ follows the χ^2-distribution. The result when k is large requires more intensive study.

P. P. N. Nayer (1936) has fitted a Type I curve to the L_1 distribution for a single variable. This approximation will be referred to as A_1. Subsequently Nair (1938) has given a formal solution to the distribution and for the case of samples of equal size the exact expression has been derived. The intrinsic accuracy of the A_1 approximation for the case when samples are of equal size has been tested by Bishop and Nair (1939) with the help of the results of the latter.

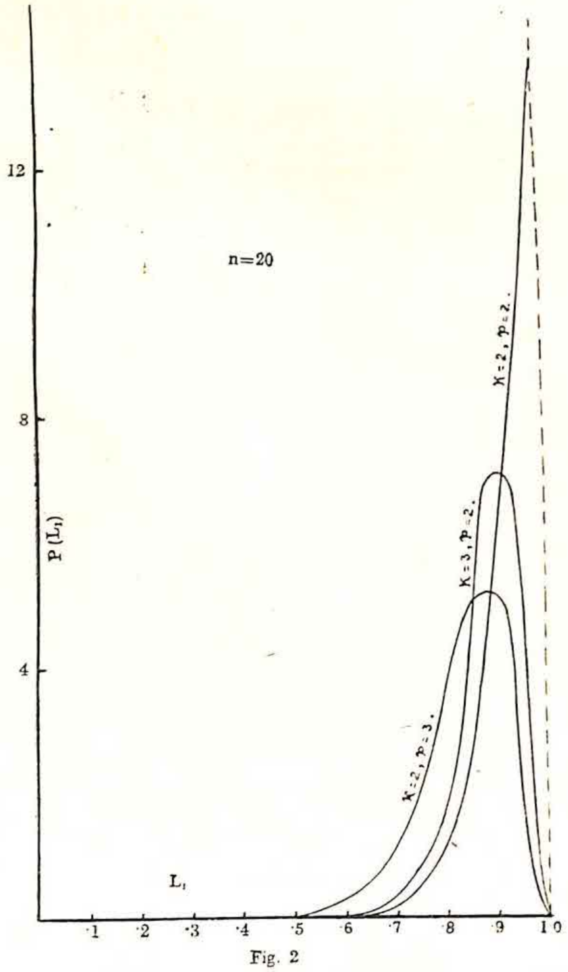


Fig. 2

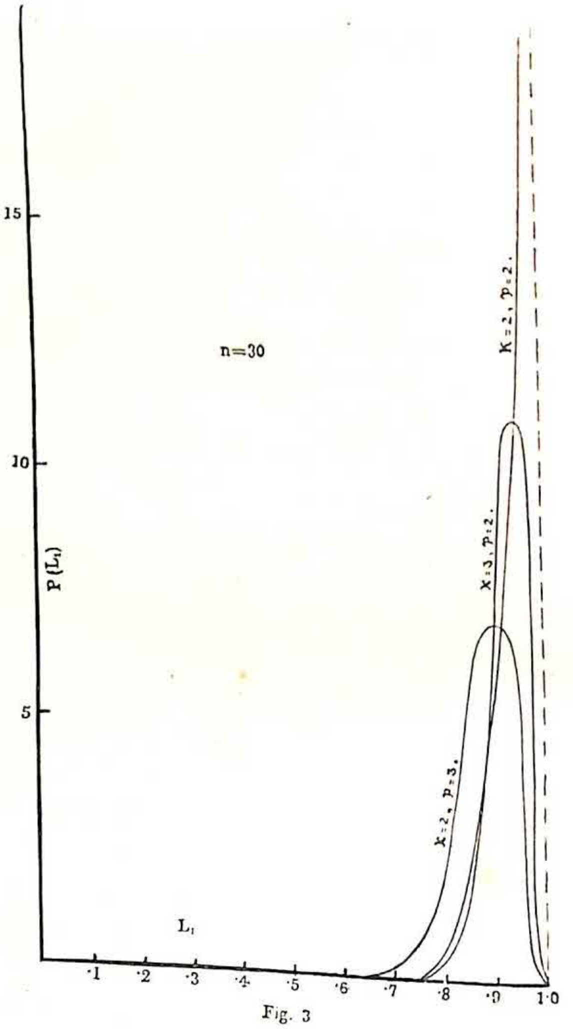


Fig. 3

Bishop (1939) has given an empirical method (A_2) of fitting a Type I curve to the λ-distribution. Bishop has also obtained an approximation (A_3) for the general case in terms of the χ^2-distribution. Bartlett (1937) has given a modified definition of the λ-criterion for a single variable and an approximate distribution (A_4) for the same. Pitman (1939) has suggested an approximation (A_5) for the distribution and verifies its accuracy for a few cases. In the present paper an attempt is made to test the accuracy of the different methods of approximation with the help of the exact distributions developed by Nair. To simplify numerical work, only certain types of samples have been taken for verification. A general method of fitting the Type I approximation is also suggested.

4. ACCURACY OF THE DIFFERENT APPROXIMATIONS

The exact distributions obtained are helpful in testing the accuracy of the different approximations for $P(L_1)$.

(a) *Approximation A_2.* Defining $\bar{L}_1 = L_1^{1/2}$*,

$$P(\bar{L}_1) = \text{const.} \bar{L}_1^{m_1-1} \times (1-\bar{L}_1)^{m_2-1}$$

where $m_1 = k(n-p) - 0.01(k-1)(90-39p+9p^2)$, $m_2 = 0{\cdot}25(k-1)\,p(p+1)$. .. (4·1)

Using Beta Function Tables the 5 per cent probability level for $\bar{L}_1$ is calculated, and since

$$P(a \leqslant \bar{L}_1 \leqslant b) = P(a^2 \leqslant \bar{L}_1^2 \leqslant b^2) = P(a^2 \leqslant L_1 \leqslant b^2) \qquad .. \ (4{\cdot}2)$$

the corresponding probability level for L_1 may be calculated. The probability for this value of L_1 is now calculated with the help of the exact distribution. The results of these calculations are given below.

TABLE 1. TRUE PROBABILITY CORRESPONDING TO THE FIVE PER CENT LEVEL FROM A_2 APPROXIMATION.

n	L_1	probability	n	L_1	probability
	$p=1$: $k=3$			$p=3$; $k=2$	
10	·7928	·0502	10	·4171	·0461
20	·8982	·0500	15	·5972	·0475
30	·9326	·0500	25	·7533	·0485
			30	·7935	·0488
			40	·8442	·0488
	$p=2$: $k=2$			$p=3$: $k=3$	
10	·6091	·0492	15	·5693	·0481
15	·7384	·0497	30	·7704	·0484
20	·8037	·0501			
25	·8430	·0503			
	$p=2$: $k=3$			$p=4$: $k=2$	
10	·5918	·0491	20	·5756	·0461
20	·7921	·0496	30	·7090	·0477
30	·8609	·0497			

Bishop has done considerable work to test the adequacy of his approximation. His methods are mostly confined to calculations of the first four moments from the true distribution of L_1 and comparing them with the values obtained from the approximation (A_2). From Table 1, one is inclined to feel that this approximation may not be satisfactory in the case of p greater than 2 even when the sample size is forty.

(b) *Approximation A_3.* Bishop has proved that in large samples $\chi^2 = -(2N/G)\log_e \bar{L}_1$ follows the χ^2 distribution with f degrees of freedom, where

$$G = 1 + \frac{1}{f}\left[\sum_{i=1}^{k}\sum_{i=1}^{p}\{i^2/2n_i + i/(n_i-i) + n_i/3(n_i-i)^2\}\right]$$
$$-\sum_{i=1}^{p}\left[(k+i-1)^2/2N + (k+i-1)/(N-k+1-i) + N/3(N-k+1-i)^2\right] \qquad .. \ (4{\cdot}3)$$

and $f = 0{\cdot}5(k-1)p(p+1)$. The accuracy of this approximation has been tested with the help of the true distribution and the results are given in table 2.

---

* L_1 is the same as Bishop's 'l_1'

TABLE 2. TRUE PROBABILITY CORRESPONDING TO ONE PER CENT AND FIVE PER CENT LEVELS FROM A_3 APPROXIMATION

n	L_1 (·01)	probability	L_1 (·05)	probability
(A) Equal Samples				
$p = 2 ; k = 2$				
10	·4920	·0105	·6134	·0518
20	·7293	·0102	·8045	·0509
30	·8164	·0102	·8695	·0508
$p = 2 ; k = 3$				
10	·5009	·0106	·5958	·0521
20	·7334	·0102	·7928	·0505
30	·8190	0101	·8611	·0501
$p = 3 ; k = 2$				
10	·3274	·0120	·4333	·0549
15	·5103	·0108	·6042	·0528
20	·6181	·0105	·6974	·0510
30	·7362	·0102	·7950	·0506
$p = 3 ; k = 3$				
10	·3208	·0124	·4018	·0578
15	·5014	·0110	·5749	·0537
20	·6093	·0106	·6721	·0508
30	·7287	·0102	·7758	·0500
$p = 4 ; k = 2$				
20	·5030	·0106	·5816	·0514
30	·6490	·0100	·7110	·0501
(B) Unequal Samples				
$p = 2 ; k = 3, (n_1 = n_2 = n, n_3 = 2n)$				
5	·2828	·0121	·3883	·0578
10	·6034	·0106	·6850	·0519
15	·7292	·0103	·7893	·0508
20	·7948	·0101	·8420	·0502
$p = 2 ; k = 3, (n_1 = n, n_2 = 2n, n_3 = 3n)$				
5	·4673	·0123	·5656	·0573
10	·7247	·0105	·7857	·0517
15	·8149	·0102	·8579	·0506
$p = 3 ; k = 2, (n_1 = n, n_2 = 2n)$				
15	·6484	·0106	·6147	·0518
20	·7315	·0103	·7912	·0501

From Tables 1 and 2 it is evident that the A_3 approximation is better than A_2. When the number of variables increases tests of significance will be valid only if the size of the sample is also increased. With this remark, A_3 approximation seems fairly good even for high values of p, the sample size being made correspondingly large.

(c) *Approximaiions* A_4 *and* A_5. These approximations refer only to $p=1$, Bartlett defines

$$\mu^2 = \prod_{t=1}^{k}\left[V_t^{f_t}\right] \Big/ \left[\left(\sum_1^k f_t V_t\right) \Big/ F\right]^F \qquad .. \quad (4{\cdot}4)$$

where V_t $(t = 1, 2, \ldots, k)$ are k independent variances with f_t degrees of freedom, taken as estimates of a common σ^2. When the f's are equal and $F=kf$, $\mu^2=L_1^F$, that are mation A_4 is that $-(2 \log_e\mu)/c$ where $c = 1 + [1/f_t - 1/F]/3(k-1)$ is distributed like χ^2 with Approxi- $(k-1)$ degrees of freedom. Pitman defines L by the relation $\mu=\exp.(-L)$ and shows that $2L/(1+\alpha)$ where $(1+\alpha)=c$ follows the χ^2 distribution with $(k-1)$ degrees of freedom. Thus Bartlett's and Pitman's approximations are practically the same.

The substantial accuracy of this approximation has been studied by Pitman for a few cases with the help of elliptic integrals. To test the accuracy of this approximation the distribution of L_1 in (2.2) cannot be used. Hence defining $\bar{L} = \exp.(L/F)$, the distribution of $\bar{L}$ is derived. It can be easily proved that the moment function of $\bar{L}$ is

$$\varphi(t)=\frac{F^{t/2}\,\Gamma(F/2)}{\Gamma(F/2+t/2)}\cdot\prod_1^k\frac{\Gamma(f_i/2+tf_i/2F)}{\Gamma(f_i/2)f_i^{tf_i/2F}} \qquad .. \quad (4{\cdot}5)$$

Hence using Mellin's transformation

$$P(\bar{L}) = \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} (\bar{L})^{-t-1} . \phi)t)dt \qquad \ldots (4{\cdot}51)$$

This integral has been evaluated for the following cases :

$k=3$, (1) $f_1=f_2=f,\ f_3=2f$ and $F=4f$

(2) $f_1=f,\ f_2=2f,\ f_3=3f$ and $F=6f$

The following table gives the accuracy of approximation involved in the A_5 test in the cases mentioned above where the exact distribution of $\bar{L}$ has been derived.

TABLE 3. TRUE PROBABILITY CORRESPONDING TO ONE PER CENT AND FIVE PER CENT LEVELS FROM A_5 APPROXIMATION

f	$L_1(\cdot 01)$	probability	$L_2(\cdot 05)$	probability
	($k=2, f_1=f_2=f, f_3=2f$)			
1	·2054	·0053	·3571	·0371
2	·5048	·0084	·6411	·0467
3	·6494	·0093	·7552	·0493
4	·7299	·0096	·8184	·0493
6	·8156	·0098	·8758	·0497
10	·8874	·0099	·9252	·0499
20	·9430	·0100	·9626	·0499
	($k=3, f_1=f, f_2=2f, f_3=3f$)			
1	·3751	·0068	·5284	·0419
2	·6459	·0090	·7525	·0481
3	·7561	·0096	·8337	·0493
4	·8145	·0098	·8750	·0496
5	·8504	·0099	·9000	·0496
6	·8747	·0099	·9166	·0499
10	·9241	·0100	·9500	·0499
20	·9618	·0100	·9750	·0499

It is evident that the approximation is fairly good except for very small values of f.

5. TYPE I APPROXIMATION OF $\bar{L}_1$

Bishop's approximation A_2 may be viewed mathematically by assuming that

$$P(\bar{L}_1) = \text{const. } \bar{L}_1^{\,N-kp+\mu-1}\,(1-\bar{L}_1)^{f/2-1} \qquad \ldots (5{\cdot}1)$$

The moment function of $\bar{L}_1$ from this approximation is

$$\phi(t) = \frac{\Gamma(N-kp+\mu+t).\ \Gamma(N-kp+\mu+f/2)}{\Gamma(N-kp+\mu+t+f/2).\ \Gamma(N-kp+\mu)} \qquad \ldots (5.2)$$

Using Stirling's approximation the development of (5.2) gives

$$\text{Log}_e\,\phi(t) \sim -\frac{ft}{2N} + \frac{\frac{ft^2}{4} + \left(\mu - kp + \frac{f}{4} - \frac{1}{2}\right)\frac{ft}{2}}{N^2} \qquad \ldots (5{\cdot}3)$$

and the corresponding asymptotic expansion of the true moment is

$$\text{Log}_e\,\phi(t) \sim -\frac{(k-1)p(p+1)t}{4N} + \frac{t^2(k-1)p(p+1)}{8N^2}$$

$$-\frac{[(2p^2+9p+11)\ \Sigma\ 1/n_\beta - (2p^2+3p-1) + 6k(k+p+1)/N]tp}{24N}$$

Comparing the expansions (5.3) and (5.4) it is seen that

$$f=0.5(k-1)p(p+1)$$

and

$$\mu=kp-\frac{p(p+1)\,(k-1)-4}{8}+\frac{k+2+p}{p+1}-\frac{2p^2+9p+11}{6(k-1)\,(p+1)}\left(N\Sigma\frac{1}{n_\beta}-1\right)\,.. \quad (5\cdot5)$$

This will give very good agreement between the moments calculated from the true and approximate distributions.

CONCLUSION

The main object of this paper has been to examine the accuracy of the levels of significance for the L_1 criterion as calculated from the different approximations. The approximation for $-\log_e L_1$ with the correction factor seems the best from the point of view of giving correct levels. The Type I approximation seems definitely inferior for $p\geqslant3$.

The author wishes to express her deep gratitude to Dr. U. S. Nair for the valuable help and guidance she had in the preparation of this paper. She is also thankful to the University of Travancore for the encouragement given in carrying out this work.

REFERENCES

1. BARTLETT. (1937) Properties of sufficiency and statistical tests. *Proc. Roy. Soc.*, **160**(A), 268-282

2. BISHOP. (1939) On a comprehensive test for the homogeneity of variances and covariances in multivariate problems. *Biometrika*, **31**, 31-35.

3. BISHOP AND NAIR (1939) A note on certain methods of testing for the homogeneity of a set of estimated variances. *Jour. R, Statist. Soc. Supp.*, **6**, 89-99.

4. HARTLEY. (1940) Testing the homogeneity of a set of variances. *Biometrika*, **31** (3 & 4), 249-255.

5. NAIR, U. S. (1938) The application of the moment function in the study of distribution laws in statistics. *Biometrika*, **30**, 274-294.

6. NAYER, P. P. N. (1936) An investigation into the application of Neyman and Pearson's L_1 tests, with tables of percentage limits. *Statist. Res. Mem.*, **1**, 38-51.

7. NEYMAN, J. AND PEARSON, E. S. (1930) On the problem of two samples. *Bulletin de L' Academic Polonaise des Sciences et des Letters.* Series. **A**, 471-494.

8. NEYMAN, J. AND PEARSON, E. S. (1931) On the problem of *K* samples. *Bulletin de L' Academic Polonaise des Sciences et des Letters.* Serie **A**, 460-481.

9. PITMAN. (1939) Tests of hypotheses concerning location and scale parameters. *Biometrika*, **31**, (1 & 2), 200-215.

10. WILKS, S. S. (1932) Certain generalisations in the analysis of variance. *Biometrika*, **24**, 471-494.

Paper received : 15 November, 1943.

A NOTE ON THE DISTRIBUTION OF THE SUM OF CHI-SQUARES

By A. BHATTACHARYYA
Statistical Laboratory, Calcutta

Mr. B. C. Bhattacharyya (1943) has considerd the distribution of the sum of two chi-squares, each of which is the sum of squares of the same number of independent normal variates with zero mean and the same standard deviation, but the two sets differing in their standard deviations. He has shown that the distribution involves some Bessel functions considered by Mckay. We shall here briefly indicate how the amo distribution can be expressed as a convergent series in Laguerre's polynomials. We shall consider he more general case where the two chi-squares are not independently distributed.

2. Let $(x_1, y_1;\ x_2, y_2;\ \ldots\ x_n, y_n)$ be n pairs of independent correlated normal variates with zero mean and s.d.'s σ_1 and σ_2; also let the correlation coefficient be ρ. Let us define χ^2_1 as the sum of squares of the x's and χ^2_2 that of the y's. The case when the two χ^2 are independent is obtained when $\rho=0$. We shall consider the distribution of $\chi^2_1 + \chi^2_2$. The characteristic function of $x^2_1 + y^2_1$ is

$$\phi(t) = [1 - 2it\,(\sigma^2_1 + \sigma^2_2) - 4(1 - \rho^2)\,\sigma^2_1\,\sigma^2_2\,t^2]^{-\frac{1}{2}}$$

and hence the characteristic function of $\chi^2_1+\chi^2_2$ is $[\phi(t)]^n$. From the form it takes it is obvious that the same Bessel function form is valid in this case also. Similarly it is easily seen that the distribution of $\chi^2_1-\chi^2_2$ follows the other Bessel function form (Bhattacharyya, 1943).

The characteristic function of $u = (\chi^2_1 + \chi^2_2)/(\sigma^2_1 + \sigma^2_2)$ is easily found to be

$$\psi(t) = [1 - 2it - 4t^2\,\sigma^2_1\,\sigma^2_2(1-\rho^2)\,(\sigma^2_1 + \sigma^2_2)^{-2}]^{-n/2} = [(1 - it^2)^2 + at^2]^{-n/2} \qquad ..\ (2{\cdot}1)$$

where $a = 1 - 4\sigma^2_1\,\sigma^2_2(1 - \rho^2)\,(\sigma^2_1 + \sigma^2_2)^{-2}$. As $2\sigma_1\,\sigma_2 \leqslant \sigma^2_1 + \sigma^2_2$, and $\rho^2 \leqslant 1$, we must have $|a|< 1$ and so $|at^2(1 - it)^{-2}| = |a|\,t^2(1 + t^2)^{-1} < 1$, for all values of t. So, from (2.1) we get

$$\psi(t) = (1 - it)^{-n}\,[1 + at^2(1 - it)^{-2}]^{-n/2} = \sum_{r=0}^{\infty} \binom{-n/2}{r} a^r\,t^{2r}\,(1-it)^{-n-2r} \qquad ..\ (2{\cdot}2)$$

The infinite series is uniformly convergent for all values of t. Now, differentiating the relation

$$\int_{-\infty}^{\infty} (1 - it)^{-n-2r}\,\text{Exp.}(-\,it\,u)dt/2\pi = e^{-u}\,u^{n+2r-1}/\Gamma(n + 2r)$$

r-times with respect to u, we get

$$\int_{-\infty}^{\infty} (1 - it)^{-n-2r}\,(-\,it)^{2r}\,\text{Exp.}(-\,it\,u)dt/2\pi = \left(\frac{d}{du}\right)^{2r} e^{-u}\,u^{n+2r-1}\,/\Gamma(n+2r)$$

$$= \Gamma(n)\,e^{-u}\,u^{n-1}\,\mathrm{L}_{2r}\,(u, n)/\Gamma(n + 2r) \qquad ..\ (2{\cdot}3)$$

where $\mathrm{L}_{2r}\,(u, n)$ is a Laguerre's polynomial. Hence applying Fourier's integral theorem to $\psi(t)$ and using (2.3) we get the distribution function of u as

$$\Gamma(n)\,e^{-u}\,u^{n-1} \sum_{r=0}^{\infty} \binom{-\,n/2}{r} a^r\,\mathrm{L}_{2r}(u, n)/\Gamma(2r + n) \qquad ..\ (2{\cdot}4)$$

The above series is convergent for all values of u. Proceeding in a similar way it could be shown, under some restrictions, that the distribution of the sum of three or more χ^2's can be expressed in a series o Laguerre's polynomials.

3. The non-null distribution of χ^2 can be expressed, under certain conditions, in a series of Laguerre's polynomials. Let x_k, $(k = 1, 2, \ldots n)$ be n independent normal variates with zero mean and standard deviation σ_k. Then the distribution of $\chi^2 = \sum\limits_{k=1}^{n} x^2_k$ can be regarded as the non-null distribution. There is another type of non-nullity, namely when the normal variates have the same standard deviation, but different means. This has already been considered by Wicksell (1933). The classical D^2-statistic follows the same distribution (Bose, 1936). We shall consider the other type of non-null case. The characteristic function of χ^2 is obviously

$$\prod_{k=1}^{n} (1 - 2it\, \sigma_k{}^2)^{-\frac{1}{2}} \qquad .. \quad (3{\cdot}1)$$

If $n\sigma^2 = \sum\limits_{k=1}^{n} \sigma^2_k$ and $\varepsilon_k = \sigma^2_k - \sigma^2$ then the above characteristic function can be written as

$$\prod_{k=1}^{n} (1 - 2it\, \sigma^2 - 2i\, \varepsilon_k\, t)^{-\frac{1}{2}} = (1 - 2it\, \sigma^2)^{-n/2} \prod_{k=1} \{1 - 2it\, \varepsilon_k/(1 - 2it\, \sigma^2)\}^{-\frac{1}{2}}$$

The characteristic function of χ^2/σ^2 is

$$(1 - 2it)^{-n/2} \prod_{k=1}^{n} \{1 - 2it\, \varepsilon_k/\sigma^2(1 - 2it)\}^{-\frac{1}{2}} \qquad .. \quad (3{\cdot}2)$$

If now we suppose that $\varepsilon^2_k < \sigma^4$ for all values of k, then each of the expressions $\{1 - 2it\, \varepsilon_k/\sigma^2(1 - 2it)\}^{-\frac{1}{2}}$ can be expanded in a uniformly convergent series, which then can be multiplied together giving a uniformly convergent infinite series for (3.2) of the form

$$\sum_{r=0}^{\infty} c_r (it)^r\, (1 - 2it)^{-r-n/2} \qquad .. \quad (3{\cdot}3)$$

where $c_0 = 1$, $c_1 = 0$, $c_2 = \Sigma\, \varepsilon^2_k/\sigma^4$, $c_3 = 4\Sigma\, \varepsilon^3_k/3\sigma^6$ etc. Applying, as in section 2, Fourier'a integral theorem and remembering (2.3) we get for the distribution function of $x = \chi^2/\sigma^2$ an expression of the form

$$\Gamma(n/2)\, e^{-x/2}\, x^{n-1} \sum_{r=0}^{\infty} d_r\, L_r(x, n)/\Gamma(n + r)$$

where of course d_r depends on c_r or ultimately on ε_r.

There are many other non-null forms of χ^2 which we cannot adequately consider in this short note. It is hoped to consider these in detail at some future date, when other interesting properties of χ^2 will be considered.

REFERENCES

1. Bhattacharyya, B. C. (1943) On an aspect of pearsonian system of curves and few analogies. *Sankhya*, **6** (4), p. 415-418.

2. Bose, R. C. (1936) On the exact distribution and moment-coefficients of the D^2 -statistic. *Sankhya*, **2** (2), p. 143-154.

3. Wicksell, S. D. (1933) On correlation functions of type III. *Biometrika*, **25**, 121-133.

Paper received : 29 September, 1943.

REPORT ON THE BIHAR CROP SURVEY: RABI SEASON 1943-44*

BY P. C. MAHALANOBIS

Statistical Laboratory, Calcutta

CONTENTS

PAGES

PART 1. GENERAL DESCRIPTION OF THE SURVEY.

Section 1. General Introduction .. .. 29—32
Section 2. Method of the Sample Survey .. .. 32—38
Section 3. The Design of the Survey .. .. 38—41
Section 4. General Account of the Survey .. .. 41—45

PART 2. RESULTS OF THE RABI SURVEY.

Section 5. Area under Rabi Crops 1943-44 .. .. 46—53
Section 6. Rate of Yield and Outturn of Wheat and Gram 1943-44 .. 53—63
Section 7. The Problem of Mixed Crops .. .. 63—68
Section 8. Cost of Operations .. .. 68—76
Section 9. Accuracy of Area Enumeration .. .. 76—87

PART 3. GENERAL REVIEW AND SUMMARY.

Section 10. General Review of Methods for the Improvement of Crop Statistics .. .. 88—99
Section 11. Summary of the Report .. .. 99—106

PART 1. GENERAL DESCRIPTION OF THE SURVEY

SECTION 1. GENERAL INTRODUCTION

HISTORY OF THE SCHEME

First enquiry. The Government of Bihar in letter No. 13297 P. C. dated 28th September 1943 enquired whether I could undertake as early as possible a sample survey of rice in Bihar on the same lines as the survey of the jute crop in Bengal which had been going on under my guidance for several years. In my reply dated 4th October I pointed out that the season was already too far advanced for an area survey of *aghani* (winter paddy) but crop-cutting work of an exploratory nature might possibly be still undertaken. Mr. R. A. E. Williams, I.C.S. (then Secretary, Revenue Department) in his letter dated 27th October 1943 stated that Government would like me to organize crop-cutting work on paddy and requested me to submit a scheme with estimates. By this time it had become too late for even crop-cutting work on paddy. After some further correspondence it was decided that I should go to Patna for personal discussions.

2. *Discussion in Patna.* On the 16th November 1943 I reached Patna and in the afternoon of the same day had a three-hour conversation with Mr. R. A. E. Williams. In the morning of the 17th November I had the opportunity of discussing the matter personally with His Excellency Mr. (now Sir) Francis Mudie, Governor of Bihar, who after examining my proposals approved of them generally and asked me to submit a scheme as soon as possible. On the 18th and 19th November I had further talks with a number of Bihar officials, collected a good deal of material in the form of Settlement Reports, and made personal enquiries in the Survey Office about Cadastral Survey village maps.

3. *A five-year scheme.* On the basis of the information collected in this way I prepared, with the assistance of Mr. J. M. Sengupta, Statistician, who had accompanied me to Patna, a five-year scheme for the improvement of crop statistics in Bihar by the

---

* This Report was submitted to the Government of Bihar on 5th November, 1944.

method of the sample survey at an estimated cost of twenty-four lakhs of rupees exclusive of dearness allowance, and submitted this scheme at Patna on 19th November 1943. I explained that the objects of the scheme would be :—

(a) to develop an efficient sampling technique for estimating acreage, yield per acre, and total outturn of important crops such as paddy, wheat, barley, maize, gram, etc. and the total cropped area in each of the three (*bhadoi*, *aghani* and *rabi*) seasons in Bihar; and secondly,

(b) to build up the necessary human organisation on both the statistical and field sides of the work for carrying out the sample survey on standardized lines from the fifth year of the scheme.

4. The approach in a scheme of this kind must be definitely experimental. I, therefore, proposed starting with a pilot or exploratory survey on a small scale in the *rabi* season of 1943-44, and gradually expanding the scale of operations working up to a full provincial survey in the *aghani* (winter) season of 1945-46. The scheme visualized a recurring expenditure of five lakhs of rupees per year for work in three crop seasons: *bhadoi* (autumn), *aghani* (winter), and *rabi* (spring). The field work would be done under the direct control of the Government of Bihar, but the statistical and technical work was to be done in the beginning by the Indian Statistical Institute under my guidance as honorary Statistical Adviser, and was to be gradually transferred from the third year to the Government of Bihar who would take over the statistical section completely at the end of the five year scheme. A large portion of the Statistical Laboratory had been removed from Calcutta to Giridih early in 1942 as an evacuation measure; and it was suggested that the bulk of the statistical work in the Bihar crop survey scheme should be located at Giridih.

5. *Beginning of preparatory work.* The scheme submitted by me was approved in principle by the Government of Bihar within a few days, and I was asked to make necessary arrangements for the survey in the *rabi* season of 1943-44. Orders were issued by Government on the 30th November 1943 requesting all District Collectors to send two complete sets of district, *thana*, and C.S. village maps to the Giridih Branch of the Indian Statistical Institute. In his letter No. 28/1c-1/14 dated 4th January 1944 Mr. Williams assured us that Government would give the Institute full support and facilities for the work.

6. *First plan abandoned.* Government was keen that we should cover the important wheat districts in the province, and a design of the proposed survey was prepared on this basis. Unfortunately preparatory work could not be started as necessary village maps had not been received by the middle of December. It was clear that the preparatory work could not possibly be completed in time, and there was no other alternative but to abandon the original plan.

7. *Revised plan for the rabi survey.* A second series of meetings and conferences was therefore arranged at Patna on the 17th and 18th December1943 in the course of which I met Mr. E. R. J. R. Cousins, I. C. S. (Adviser to H. E. the Governor), Mr. A. P. Middleton, I. C. S. (Member, Board of Revenue), Mr.R. A. E. Williams, I. C. S. (Secretary, Revenue Department) and a number of other officials. I also had a full discussion with Rai Bahadur Sinheswar Prasad Sinha, a senior Deputy Collector, who had been appointed to the newly created post of whole-time Superintendent of Statistical Surveys. As a result of these discussions it was decided to carry out an exploratory survey in the *rabi* season of 1943-44 in the two districts Monghyr and Shahabad for which village maps had been already received.

8. *Progress of the survey.* The preparatory work for this survey was begun in the Statistical Laboratory at Giridih at the end of December 1943 and was completed after strenuous labour in the first week of February 1944. Seventy-eight men with experience of field work in Bengal were sent to Bihar on loan and 102 workers were locally recruited for the field survey which started on the 12th February and was completed on the 13th April, 1944. Crop-cutting work on wheat and gram started in the third week of March and was closed along with the area survey on the 13th April 1944. I went to Patna early in January 1944, and again at the end of March. I met Mr. Middleton and Mr Williams on the 23rd March; and on the 24th March I had an interview with His Excellency Mr. Mudie who examined with keen interest the preliminary results based on the material collected during the first fortnight of the survey.

9. *Surveys in the bhadoi and aghani seasons.* Although the present report is concerned with the *rabi* season of 1943-44 a brief mention may be made here of subsequent work. In the original programme it had been intended to organize the work on an expanding scale in the *bhadoi* (autumn) and *aghani* (winter) seasons of 1944-45. As a result of discussions in Patna at the end of March it was decided to have a survey of the early *makai* (maize) crop in certain portions of the districts of Purnea, Bhagalpur, and Monghyr. Not only this; the great success with which the field work in the *rabi* season had been organized gave us confidence in suggesting that a full provincial survey should be attempted in the *bhadoi* (autumn) season of this year (instead of in the *aghani* (winter) season of 1945-46 as contemplated in the original scheme),

and thereafter continue the provincial surveys in succeeding crop seasons. This proposal to expedite the work received the full support of H. E. Mr. Mudie, and Government sanction was given accordingly. The field work for the survey in the *bhadoi* season was started on the 3rd July and was completed on the 3rd October 1944. In quick succession the field work for the survey of *aghani* (winter) crops was begun on the 10th October and was in progress at the time of writing this report.

THE PRESENT REPORT

10. *Contents of the present report.* The present report gives an account of the survey in the *rabi* season and the results obtained from the material collected during the survey. The report consists of 11 sections arranged in three parts. Part 1 is mainly descriptive. The present (first) section gives a brief history of the scheme and a general introduction. Section 2 gives a general and non-technical description of the method of the sample survey of crops as developed in the course of the work in Bengal during the last 7 or 8 years. This is followed by a description of the details of the design or plan of the survey of *rabi* crops in Section 3 ; and a general account of the progress of the survey in Section 4.

11. Part 2 gives the actual results of the survey. Section 5 deals with the estimated acreages under five important crops, wheat, barley, gram, *arhar* and *khesari*. Section 6 is concerned with the rates of yield per acre of wheat and gram based on crop-cutting work done on these two crops. An attempt has also been made in this section to estimate the total outturn of these two crops. Comparison with official figures was not possible owing to difficulties in the interpretation of acreage estimates in the case of mixed crops and the unsatisfactory nature of official figures relating to outturn. A detailed discussion of the problem of mixed crops is given in Section 7. This is followed by a discussion of the cost of operations in Section 8. Finally, mistakes in preparing crop records in the method of complete enumeration are examined in considerable detail in Section 9.

12. In Part 3, Section 10 gives a general review of methods for the improvement of crop statistics with special reference to the permanently settled provinces like Bengal and Bihar. The respective advantages and disadvantages of the method of (*a*) complete enumeration and (*b*) sample survey have been considered in some detail in this section, and a general description has also been given of the Bengal scheme for complete enumeration which is scheduled to come into effect from the current *aman* paddy or *aghani* season of 1944-45. A combination of complete enumeration and sample survey is considered to offer the most promising line of advance in the immediate future. A summary of the Report is given in Section 11 of Part 3.

13. For convenience of reference, tables have been numbered separately by Sections, and inserted in the text where possible.

14. It will be noticed that the factual material is given in considerable detail in Part 2, Sections (5)—(9). This is the portion (together with the description of the design of the survey in Section 3) to be consulted by those who are primarily interested in the statistical details. Those who are interested in a more general way in the survey would find a description of the sampling method in Section 2, and an account of the survey itself in Section 4, and a summary of the results in Section 11.

ARRANGEMENTS WITH THE INDIAN STATISTICAL INSTITUTE

15. In the Bengal Crop Survey Scheme both the field and the statistical work is done by the Indian Statistical Institute for which a block grant is given by the Government of Bengal. The Institute has authority to make re-appropriations within the limits of the block grant and also to make all necessary arrangements for the work. Expenditure on field work and other charges incurred exclusively for the crop survey scheme are debited directly while overhead and joint charges are allocated on the basis of total volume of work ; and accounts audited by chartered accountants are submitted to Government.

16. *Block grant.* In the Bihar Crop Survey Scheme arrangements were made, at my suggestion, to have the field work done under the direct control of Government and the Board of Revenue was placed in charge of the scheme. The statistical work was entrusted to the Indian Statistical Institute, and a block grant was sanctioned for this purpose on the same conditions as in the case of the Bengal scheme.

17. *Contribution in lieu of the services of the Statistical Adviser.* The Government of Bihar had offered to pay a suitable remuneration to me for my work as Statistical Adviser. At my request I was, however, permitted to do this work on a purely honorary basis, but a

contribution of Rs. 15,000 per year to the Indian Statistical Institute was sanctioned with effect from 16th November 1943 in lieu of my services for which we offer our sincere thanks to the Government of Bihar.

ACKNOWLEDGEMENTS

18. The present scheme owed its inception to Sir Francis Mudie when he was Governor of Bihar; and the personal interest taken by him in this matter was a great encouragement to us.

19. Mr. A. P. Middleton, I.C.S. (Member, Board of Revenue) and Mr. R. A. E. Williams, I.C.S. (Secretary, Revenue Department) were largely responsible for the smooth progress of the scheme in the initial stage. The organization of the field work during the *rabi* season was, as already noted, successful beyond expectation for which credit is due to Rai Bahadur S. P. Sinha (Superintendent of Statistical Surveys). Mr. Pritiranjan Sen (Registrar, Revenue Department) helped us by promptly attending to our many requirements.

20. In submitting this Report I wish to place on record our grateful appreciation of the help we received at every stage of the work from the officials of the Government of Bihar.

SECTION 2. METHOD OF THE SAMPLE SURVEY

THE JUTE SURVEY SCHEME IN BENGAL

21. A general description of the method of the sample survey for preparing crop estimates as used in Bengal during the last few years is given in this section. The method was developed in the course of a five-year scheme which was started in 1937 by the Indian Central Jute Committee (and was financed jointly by the Committee and the Government of Bengal) for the improvement of statistics relating to acreage and outturn of the jute crop in Bengal. On my advice, the Committee decided to explore the possibilities of the sampling method, and I was entrusted with the task of developing a suitable technique for this purpose with a view to securing acreage estimates with a margin of error not exceeding five per cent and at an expenditure of the order of about a lakh and a half of rupees per year (at 1937-38 level of costs).

22. The work was started with a very small scale exploratory survey in 1937, covering only about 200 square miles with a total expenditure of about six thousand rupees for the field work. The scale of operations was steadily expanded in 1938, 1939, and 1940 culminating in a survey on a full provincial scale in 1941. The results were considered entirely satisfactory by the Indian Central Jute Committee which recommended the continuance of the sample survey by Government in future. In 1942 the scheme was taken over by the Government of Bengal and the work was carried out under my guidance by the Indian Statistical Institute on behalf of Government. The method was extended to the paddy crop in 1943, and was further enlarged to cover all important crops in 1944.

23. Reports on the work done in each year from 1937 to 1943 were submitted by me to the Indian Central Jute Committee and the Government of Bengal; some of these have been printed. A general description of the method was given in my presidential address to the section of Mathematics and Statistics of the Indian Science Congress in January 1942; and more recently, a full account of the theory has been given in a memoir on Large Scale Sample Surveys in the *Philosophical Transactions* Vol. 231B, No. 584, of the Royal Society of London. A non-technical description of the method is given in the present section; for fuller details and underlying theory reference will naturally have to be made to the publications mentioned above.

THE SAMPLING METHOD

24. A classical example of the sampling method is sampling from an urn containing a large number of balls of, say, three different colours, red, white and blue, in different proportions. A handful of balls is drawn out of the urn. It is then possible, by counting the number of balls of each colour, to make an estimate of the proportion of balls of each colour in the urn itself.

25. In the case of land, instead of balls of different colours, we have a very large number of plots under different crops (or lying fallow or being utilized for other purposes). A simple technique (analogous to sampling from the urn) would be to pick up purely at random a suitable number of plots of land, and ascertain how many of these are under wheat, or gram, or any other crop. This would be quite easy if the plots were of equal size. All one would have to do would be to give serial numbers to the plots, and to select the sample plots with the help of a series of random numbers (of which large collections are available in a printed form). Unfortunately, this method can not be used as the size of individual plots in Bihar varies

widely from a fraction of an acre to several hundreds of acres. Selection by serial number of plots would not give each acre of land the same chance of being included in the sample; and samples drawn in this way would not be truly random or representative. It is no doubt possible to get over the difficulty arising from unequal size of plots but the procedure would be laborious and expensive.

26. *Use of grids.* It is convenient therefore to use not plots but sample units of a definite size like 4-acre, or 10-acre, or 50-acre in area. In my earliest publications on the sample survey of the jute crop in Bengal I had called this sample-unit a grid*, and this term is now commonly used in technical literature.

27. The use of grids or sample-units of finite size is also convenient for other reasons. In large scale surveys covering a whole province with an area of 60 or 70 thousand square miles, only a limited number of sample-units can be examined within available resources in time and money. For example, in Bengal in a full provincial survey we have been using something like 60 or 70 thousand grids or on an average about one grid or sample-unit per square mile. The grids or sample-units are thus widely scattered, and the time required for the investigator to move from one grid to another (which is called 'journey time') is not negligible in comparison with the time required for the physical examination of the crops grown within the grid. In this situation it would be clearly uneconomical to examine or survey a single plot at each locality. It would be obviously convenient to examine a group of plots or sample-units of comparatively large size at each locality visited by the investigators as the additional time required for this purpose would be very small. Considerations of economy thus suggest the use of sample-units of a fairly good size.

PLANNING OR PREPARING THE DESIGN OF A SURVEY

28. *Zoning.* For the organization of the field work (as well as for certain reasons of a technical nature) it is convenient to divide each district into a suitable number of sub-units or zone-cells each of which is a compact area roughly of an equal size. In Bihar in the *rabi* season we used zone-cells each of size about 50 square miles. In the sampling method the procedure then is to locate purely at random a suitable number of sample-units or grids within each zone-cell ; and have all the plots falling within each grid examined for the proportion of land under different crops by investigators who go round from one grid to another for this purpose.

29. In planning a sample survey several things have to be settled. For example, it is necessary to decide (*a*) what should be the size or area of each sample-unit or grid (which need not be same in all zone-cells and may vary from one zone to another) ; (*b*) what should be the total number of grids used in each district or in the province as a whole ; and (*c*) how should these grids be distributed among different zone-cells, that is, how many grids should be allotted to each zone-cell.

30. *Cost of operations.* Certain considerations which have to be kept in mind in planning a survey may be now explained. The total expenditure which can be incurred, and hence the total number of field investigators at our disposal, is fixed. If we use grids of a large size, say 50-acre or 100-acre each, then the time and hence the expense involved in the physical examination of each grid would be comparatively large, and we would be able to have only a comparatively small number of grids. They will be, therefore, widely scattered ; the density (or number of grids per square mile) would be small ; and comparatively greater time would be therefore spent in moving from one grid to another. On the other hand, if the grids are of a small size there can be more of them, so that they would lie closer together and would have a higher density. The time required for moving from one grid to another would, therefore, be proportionately smaller.

31. The need of working within the limits of a fixed budget places a restriction on the choice of the size (that is, the area) of each individual grid and their total number or density per square mile. Once the size (or area) of each individual grid or sample-unit is fixed, the total number which one can have also becomes fixed. On the other hand, if the total number (or density) of sample-units is fixed then the size or area of each individual sample-unit also becomes determined. It is thus not possible to choose independently both the area of each individual grid and their total number. Cost thus supplies one connexion between area and the number of sample-units. This is why a study of the 'cost function' or how the cost

---

* I may briefly explain the origin of the term. I visualised the whole district (or the province) as divided into a large number of cells by a system of meshes formed by two perpendicular sets of parallel lines, that is, I imagined the whole area to be covered by a system of grids. I considered a sample-unit to be one of the basic cells (or grids) picked up at random. This led to the term 'grid' being used as the name of a sample-unit in the survey of crop acreage. [This term had also been used in U.S.A., and is current there.]

depends on (*a*) the size of each individual grid and (*b*) their density or total number essential for preparing efficient designs.

32. *Precision of results.* But cost is not the only factor. It is also necessary to consider the precision or the margin of error of the final estimate of crop acreage. The margin of error for estimates from individual sample-units or grids would be large when the size of each sample-unit is small. The precision of the estimate based on individual grids would increase as the size of each individual grid is increased.

33. The precision of the average or final value calculated from a large number of sample-units or grids is, however, not determined only by the precision of individual grids, but also depends on the total number of grids. We have already seen that working with grids of a comparatively large size would inevitably mean having fewer of them. On the other hand, one could have a large number of grids of a smaller size. We thus find that we can use grids of a comparatively large size for each of which the precision would be large but the average value would be based on a small number of grids. Working with grids of a small size, the precision of individual grids would be small, but the average value would be based on a much larger number of grids.

34. In this situation whether there is any net gain or not in the precision of the final average value would depend on whether the decrease in precision due to a reduction in the size of each individual sample-unit is or is not compensated for by the increase in precision of the average value due to the increase in the number of sample-units on which the average is based. This question has therefore to be settled experimentally by studying how the precision of the result changes as the size of each invididual sample-unit is changed. (The precision or the margin of error is usually studied in terms of what is known as 'variance' in technical language. The study of variance as a function of the size of individual sample-units is thus necessary to settle the present question).

35. *Aim of efficient planning.* To recapitulate, the sampling technique or the design of the survey has to be settled by deciding how the whole area is to be divided into a suitable number of zones, and what would be the size of individual grids in different zones, and their total number and distribution among different zones keeping in view :

(*a*) the amount of money (and hence of labour) which can be used for the sample survey ; and

(*b*) the degree of precision which is desired to be attained in the final estimate of crop acres.

36. The aim of preparing the design of the survey may, therefore, be stated in the following form :.

(A) If the total amount of money is fixed then our object would be to settle the size of each individual grid or sample-unit and their total number (or density) and distribution among different zones in such a way as to obtain the final result with the least possible margin of error.

(B) Alternatively, if the margin of error of the final estimate is determined in advance then the size, density and distribution of grids have to be settled in such a way that the sample survey can be done at minimum cost.

PREPARATORY WORK IN THE STATISTICAL LABORATORY

37. The different stages of a sample survey can now be described. The first thing to be settled is the design of the survey (including the division of the whole area into a suitable number of zones) as explained above. Once this is done the preparatory work can start in the Statistical Laboratory.

38. *Location of sample-units or grids.* For this purpose Cadastral Survey village maps (usually on the scale 16 inches to a mile printed on sheets of about 30″×22″) are used ; the boundary of each plot together with the revenue serial number of the plot is shown in these maps. A simple apparatus (called a co-ordinatograph) was devised for locating points at random on these maps. This is essentially a frame of rectangular shape made of four pieces of wood joined at right-angles at four corners. Two wooden scales (called frame-scales) are mounted on the two opposite (and longer) sides of the co-ordinatograph ; a bridge (which is simply a piece of wood of about the same width as the frame-work) rests perpendicularly to the two sides containing the scales ; and slides freely on them. A third wooden scale (which is called the bridge-scale) is mounted on the bridge on which slides freely a cursor carrying a pencil. The bridge thus moves parallel to the two longer sides of the frame while the pencil on the bridge moves in a perpendicular direction.

39. The operational procedure is simple. The co-ordinatograph is placed on a map. A pair of random numbers say (x) and (y), each consisting of three digits, is taken from avail-

able tables. The sliding bridge has an arrow mark with a notch at either end which is adjusted on the frame-scales at the reading corresponding to the random number (x) ; the sliding pencil on the bridge is adjusted at the reading (y) on the bridge-scale, and the pencil is lowered to mark the point on the map. The co-ordinates of this point (with an arbitrary origin) are thus (x) and (y) ; and as this is a pair of random numbers, the position of the point is also purely random over the map.

40. The actual location of the points on the village map is done in three successive stages. A map is prepared on a large scale of 4 miles to the inch for districts (or 16 miles to the inch for the whole province) under survey showing the division into zone-cells. Photographs are now prepared of each zone-cell on a larger scale of about one inch to the mile showing the boundaries of each *mauza* (village) sheet of map falling within each zone-cell. These maps are placed under the co-ordinatoraph, and the required number of points (one for each grid) in accordance with the design of the survey are located at random. Each *mauza* sheet thus gets a number of points located on it at random which fixes the number of grids, it being possible, of course, for some of the sheets to get no points at all, or to get one point or more than one point each. Finally, the co-ordinatograph is placed on individual mauza sheets and the appropriate number of grids are located in the way already described.

41. In the above way the position of sample-units or grids are marked at random on appropriate village maps. Grids of the appropriate size (4-acre in the *rabi* season of 1944) are then stamped on the maps with the help of engraved brass plates. Each grid is marked in a specified manner using the sample point as the south-west corner point of the square. The grids are also serially numbered as they are marked on the village maps. This finishes what is called the location of the grids on maps.

42. *Khasra or field lists.* A list is then prepared of the revenue serial numbers of all plots which lie either entirely within or partly within and partly outside each grid. A single slip of paper is used for each grid for convenience of field survey and subsequent tabulation work. At the top of each such sheet is given the serial number of the grid, its size in acres, the name of the village, police station etc., as well as the serial number of the zone-cell, and the sample (A) or (B) to which it is allotted. Necessary blank columns are provided in each sheet for entering against each plot the anna-estimate of the proportion of land under different crops as explained below. The form also contains space for the signature of the investigator and certain other particulars. The lists are supplied to the Field Branch and duplicate copies are prepared for field and laboratory work.

FIELD SURVEY

43. Each investigator is supplied with the field lists of grids allotted to him together with C. S. maps of the villages in which these grids have fallen. Each field party usually consists of four investigators in charge of a field inspector. One such party works in a certain number of zone-cells. Each party carries its own bundle of maps and other forms and takes up their residence in a village, the place where each party spends the night being called a camp. The inspector in charge distributes the work among his investigators. Each investigator then goes out to the field with the appropriate village maps, identifies the plots (of which the revenue serial numbers are given in the field lists), and makes an actual physical examination of the crops growing on each of these plots. He then enters anna-estimates of the proportion of land under different crops in the appropriate columns in the khasra or field lists.

44. As soon as the work in a particular locality is completed the whole party shifts camp to another village. Usually two or three days are spent at each camp ; but occasional haltages of six or seven days are also made. The inspector in charge of each field party keeps in touch with the investigators by moving from one camp to another. He is responsible for the accuracy of the work done by his unit, and is required to check a certain proportion of the sample-units from time to time. Investigators and inspectors are supplied with detailed instructions covering different phases of the field work.

TABULATION OF CROP RECORDS

45. *Area extraction.* In the meantime the area of each individual plot included in the grid (and hence in the field list) is measured in the Statistical Laboratory. Photographic graticules or scales are used for this purpose ; these are simply lantern slides on which a graph paper is photographed on a scale such that each small square is 0·01 acre. This photographic scale is placed over the map and the area of that portion of each plot which is included within the boundaries of the grid is directly measured by counting the numbr of squares, and entered in a column provided for this purpose in the field list.

46. *Crop records.* The field investigators hand over the field list completed by them from time to time to the inspector in charge who then despatches the whole lot directly to

the Statistical Laboratory, Giridih. As already explained, anna-estimates of the proportion of land under different crops are given in this list in appropriate columns. After the field lists are received in the Statistical Laboratory, these anna-estimates are converted into decimal figures, and entered in columns provided for this purpose in the Laboratory copy of the field list. The area of that portion of the plot which falls within the grid has been already entered in another column as explained above ; multiplying the area figure by the estimated proportion of land under a particular crop, the actual area in acres under each crop is entered in a series of columns provided for this purpose in the Laboratory copy of the field list. Adding the figures in each column, the total area in acres under each crop for each sample-unit or grid is directly obtained. This finishes the stage of calculations of grid proportions or the crop acreage for each individual grid for each crop.

47. *Estimates of acreage.* For each crop, the grid proportions are then added, and in this way the average value is determined for each zone-cell. Multiplying this average value by the total area of the zone-cell, the crop acreage for each zone-cell is next directly calculated. Sub-totals are then built up for suitable administrative units, and the final estimate is obtained for each district (or the province) as a whole.

48. *Statistical analysis.* The tabulation of crop acreage is thus quite straightforward. A good deal of other statistical calculations also have to be made for the comparison of different sub-samples, estimating the margin of error, application of appropriate statistical tests, and for studies of cost and variance functions. It will not serve any useful purpose to go into these technical details for which reference must be made to the literature on the subject already cited.

PRECISION OF THE RESULTS OF A SAMPLE SURVEY

49. The precision of the results of a sample survey will depend on two things, namely, (*a*) the fluctuations or errors due to sampling ; and (*b*) mistakes in the primary field records (which depend on the reliability or otherwise of the field work). Although these two different types of errors occur together in practice it is convenient to consider them separately.

50. *Errors of sampling.* One great advantage of the sample survey is the fact that, provided the sample-units or grids are located truly at random, (and more than one unit is taken from each zone-cell), it is possible to calculate the margin of error of the final result on an objective basis. It is clear that the precision of the result will increase if we increase the size (area) of each individual grid or their total number. A little consideration will also show that if fluctuations in the intensity of cultivation are large then the margin of error will also be large. In fact, the size of grids remaining constant the margin of error of the final result increases with the heterogeneity or variability in the intensity of cultivation but decreases in proportion to the square root of the total number of grids. In the matter of crop acreages it is usually found that the variability increases as the extent or size of the region under survey is increased, but the increase in variability takes place much more slowly than the increase in the area itself.

51. Two important consequences follow. The margin of error is thus roughly the same when the same number of grids is used irrespective of the fact whether the grids are scattered over a small or large area. The margin of error with say 60,000 grids (of a given size say 4-acre) would be thus not widely different when all the grids are used in one district or when they are scattered over the whole province. The total area under wheat or any other crop is however much larger in the province as a whole than within any single district. The percentage margin of error would be thus much less for the province as a whole. In other words, other things remaining the same, the wider the area covered by a sample survey, the lower will be the percentage margin of error, that is, the greater will be the relative accuracy of the results. The sample survey thus becomes more and more efficient as the scale of operations is increased.

52. The second point which deserves notice is the fact that the margin of error can in theory be reduced to any desired extent by increasing the size and the total number of grids, that is, by increasing the total sampled area. But the greater the sampled area the greater will be the cost. Reducing the margin of error thus involves incurring greater expenditure. In practice there is, therefore, a limit beyond which it is usually not possible to increase the intensity of sampling as the cost becomes prohibitive.

53. *Recording mistakes.* As already noted, the precision of the results also depends on the quality of the field work. If the primary records are unreliable, no subsequent statistical treatment can be of any real use. Unfortunately, under existing conditions in Bengal and Bihar, mistakes do occur in primary records.* Some of these are more or less

---

* This subject has been further discussed in Section 9.

accidental and can be treated as integral parts of sampling fluctuations. There are, however, mistakes of an altogether different kind which arise from entries being made without proper physical examination of the grids due to gross negligence or shirking of work on the part of the investigators. Such mistakes or false entries are more dangerous as they are not amenable to statistical treament. Special precautions have to be taken and suitable statistical checks and controls have to be incorporated in the design of the survey to detect (and hence to discourage) such dishonest work.

54. *Interpenetrating sub-samples.* One type of control has proved extremely useful in Bengal. The total number of grids in each zone is divided into two equal portions say,(*A*)and (*B*). The grids allotted to sub-sample (*A*) are scattered at random all over the zone; and, in the same way, the grids allotted to sub-sample (*B*) are also scattered at random over the whole area; the two sub-samples (*A*) and (*B*) are thus completely mixed up and inter-penetrate into one another. The information for the two sub-samples is collected by two entirely different sets of field investigators who work independently and at different times in the same zone so that they never meet. The two sub-samples thus supply two independent estimates of crop acreage for the same zone. How far they are in agreement immediately furnishes a good idea about the reliability (or otherwise) of the results.

55. *Duplicate grids.* In practice, a certain proportion of the grids are also intentionally made common to both samples so that they are enumerated twice by the two different parties. A detailed plot by plot comparison of the records for such twice enumerated grids then shows how far the primary field work was reliable. The field staff knows that a certain proportion of grids are surveyed twice; but is not, of course, allowed to have any information as to which particular grids are being duplicated. This very knowledge of the existence of a system of duplicated grids acts as a check on dishonest work.

CROP-CUTTING WORK

56. *Yield per acre.* So far we have considered the area survey from which we can obtain estimates of the area under different crops. Our real object however is to estimate the total production of crops. It is ,therefore, also necessary to know the rates of yield per acre. In this matter the method of complete enumeration is, in fact, impracticable so that estimates must be made by some kind of sampling process. The general procedure for this purpose is to measure and enclose a suitable area in a field in which the crop is ready for being harvested and then collect the crop within the measured enclosure and directly weigh the crop. Such crop-cutting work is in many ways more complicated than area survey. It involves careful measurement of the area of the sample-cut, the weighing of the crop as harvested and or after drying and threshing. Not only this, permission to cut the crop has usually to be secured beforehand from the cultivator or the owner of the crop, and sometimes a suitable compensation has to be paid ; information had to be collected about the variety of the crop, irrigation, use of manures etc. for which individual cultivators have to be contacted.

57. *Need of proper randomization.* The fundamental issue, however, is the proper location of the sample-cuts. In India the method often followed is to choose what is called representative fields for this purpose. Extensive experimental studies have however demonstrated beyond dispute that the attempt to select representative fields by visual examination or in other ways invariably leads to biased results. F.Yates discussed this question thoroughly in his paper on 'Some examples of biased sampling.'. (*Annals of Eugenics,* Vol.VI, 1934-35 pp.202-213), and stated :-

> 'In view of the evidence, therefore, I do not think we can escape from the conclusion that the only satisfactory method of avoiding bias is for the sampling to be random, whether it be from a wheat field or from a human population.'(*p*.213.

He has also clearly stated that :—

> 'If bias is to be avoided, the selection of the samples must be determined by some process uninfluenced by the qualities of the objects sampled and free from any element of choice on the part of the observer'. (p. 202).

58. *Design of crop cutting work.* The essential requirement in crop-cutting work therefore is to prepare a design based on the principle of random sampling. In doing this it is necessary to specify (*a*) the size of each sample-cut, (*b*) the number of plots from which such sample-cuts are to be selected in each zone, and (*c*) the number and manner in which the sample-cuts are to be collected from each selected plot. Once the general design of the survey is settled, a suitable number of plots (from among those falling within grids used for the area survey) are selected beforehand at random in the Statistical Laboratory, and lists of such plots are handed over to the field staff. The investigators are then instructed to collect the necessary

number of sample-cuts from the selected plots in a specified manner which has been explained in greater detail in Section 6.

59. *Multi-stage sampling.* Certain special considerations also enter in preparing the design of a sample survey depending on the human agency which is available for the work. For example, if the number of investigators is comparatively small they will have to move about a great deal from place to place for collecting the samples. In this case it is usually necessary to prepare the design according to, what is called in technical language, the multi-stage plan. For example, a certain number of zone-cells are first selected at random; within each selected zone-cell a number of villages are next selected at random; within each such village a certain number of plots are again selected at random; and finally, within each such plot the sample-cut is located at random according to specifications. These are however details of technical procedure which need not be discussed here.

60. There is one peculiar difficulty in crop-cutting work which may be mentioned in this connexion. In the area survey it does not matter at all whether some or even all the plots included within a grid have no crop growing on them; this will be simply entered as having 'no crop'. A plot which has no crop growing on it (or the crop on which has been already harvested) is however quite useless for crop-cutting purposes. In selecting plots for such work, care has therefore to be taken to give a much larger number of plots than that from which the cuts would be actually collected in practice. Each investigator is instructed to go through the list in serial order, plot by plot, and stop as soon as the required number of sample-cuts have been collected. In case a plot is omitted and no crop is harvested from it, he is required to enter the reason for the omission, such as 'no crop growing', or 'crop not yet mature', or 'crop already harvested' as the case may be. In allowing plots to be omitted there is, however, risk of some bias coming in, but this cannot be avoided when the work has to be done by a comparatively small number of investigators who have to move about rapidly from place to place.

61. *Efficiency of sample surveys.* Broadly speaking, whether in the crop acreage or the crop yield survey, the essential feature of the sampling method is to select a suitable number of sample-units located purely at random over the whole region under survey. Estimates are then prepared on the basis of the information collected for the sample-units or grids. In practice, the total area actually sampled or surveyed is usually small; in acreage survey (as carried out in Bengal) it is of the order of half of one per cent or even less of the total area under survey. It has been found however that, with a proper design, even such a small fraction of sampled area can furnish final estimates of crop acreage with a margin of error not exceeding 3 or 4 per cent and this at a small fraction of the cost of a complete enumeration or plot by plot examination of the whole area. The respective advantages of the sample survey and of the complete enumeration have been discussed in Section 10.

SECTION 3. THE DESIGN OF THE SURVEY IN THE RABI SEASON : 1943-44

62. A general description of the method of the sample survey has been already given in Section 2. It will be remembered that the total budget being fixed, the object aimed at in preparing the design of a survey was to decide the size (that is, area) as well as the density of sample-units or grids in different regions or zones of the area under survey in such a way that the final results should have the highest possible precision.

63. In order to prepare an efficient design or programme of survey for this purpose it is necessary to have a good deal of preliminary information relating to the intensity of cultivation, that is, the proportion of land under different crops in different areas; variations in such intensity of cultivation from region to region; and the cost of enumeration for different sizes and densities of sample units etc.

64. Practically no information was available on the above points (except such divisional or district crop acreage figures as were given in old Settlement Reports). We had, however, a good deal of experience of sample survey work in Bengal extending over several years. We decided therefore to use the Bengal information in preparing the design for the first exploratory survey in Bihar making such changes as were considered necessary for the wheat crop in particular in the two districts Shahabad and Monghyr. As the field work would have to be started at the latest in the first week of February 1944 there was very little time to complete the preparatory work in the Statistical Laboratory. It was decided, therefore, to use a comparatively simple design avoiding all intricacies and complications.

65. *Size of grids.* The first point to be settled was the size (that is, area) of each sample-unit or grid. In general this size need not be constant in all regions but may vary from region

to region depending on intensity and variablity of conditions of cultivation. In order to avoid complications in the field survey in the very first season of work in Bihar it was decided, however, to use grids of only one size; and in the light of Bengal experience a grid of 4-acre of square shape was adopted.

66. *Density of grids.* The second point to be settled was the average density or number of grids per square mile. Here again in order to simplify the field work it was decided to avoid large variations and the average density to be aimed at in the exploratory stage was fixed at two grids or sample-units per square mile.

67. *Zoning.* I may now refer briefly to the question of zoning. Theoretical considerations show that if any particular area (in this case a district) is heterogeneous in the sense that the intensity of cultivation varies appreciably from one part to another then it is advantageous to divide the whole area into a suitable number of zones, the aim being to make each zone as homogeneous as possible. In Bengal we had found that the *thana* (or police station) was quite a convenient unit for purposes of zoning. In Monghyr and Shahabad we found however that the area of the revenue *thana* was rather large. It was therefore decided to split up each revenue *thana* into a number of smaller zone-cells roughly of the order of 40 or 50 square miles in area. In fact each of the districts was partitioned into 96 such zone-cells, the boundaries of which are shown in the accompanying Charts (1) and (2).

68. In partitioning these zone-cells advantage was taken of natural features like rivers or hills, but care was taken to keep the boundaries of villages intact so that the whole of the area in each village would lie entirely within a single zone-cell. The total number of sample-units to be allotted to each zone-cell was decided on the basis of an average density of two grids per square mile. The number of grids falling to the share of each zone-cell was thus simply twice its area expressed in square miles.

69. *Interpenetrating sub-samples.* The arrangement of sub-sampling has been already explained in Section 2. In this method, the grids allotted to a region are divided into two equal inter-penetrating portions or sub-samples* called sample (A) and sample (B) respectively. The distribution of the two sub-samples can be made in various ways. For example, suppose we are allotting 4,000 sample-units to a district, we may prepare the design with 2,000 sample-units and call this sub-sample (A). We may then independently repeat the whole process over again with another 2000 grids, and thus obtain a second sub-sample (B). Another method would be to locate a pair of grids at a time and then by tossing a coin (or by some other equivalent method) allot at random one of each pair to sub-sample (A) and the other to sub-sample (B). It is clear that the preparation of the sub-samples by either of the above methods would require a good deal of time. Secondly, such inter-penetration of the two sub-samples in detail would require each of the two parties of investigators to go over the entire field. Detailed inter-penetration would thus not only require a good deal of additional preparatory work but would also increase the complexity of the field survey.

70. *Sub-samples on basis of zone-cells.* For lack of time in the preparatory stage as well as for simplifying the field work we did not think it advisable to use in the *rabi* season in Bihar such detailed or fine-grained inter-penetration of the sub-samples. Instead of detailed inter-penetration of grids we arranged the 96 zone-cells into 48 pairs of adjoining zone-cells. Each pair of adjoining zone-cells comprising about 100 sq. miles was alloted at random to either sub-sample (A) or to sub-sample (B). In the method of detailed inter-penetration, work would have been done in each zone-cell by both the parties. In the present design, work in each zone-cell was done by only one party. With detailed inter-penetration a comparison between the two sub-samples would have been possible for each zone-cell separately. In the present design, valid comparison can only be made for each sub-block or for the district as a whole.

71. In the present design we thus used the method of sub-sampling in a rather large grained fashion with inter-penetration at the stage of zone-cells instead of at the stage of grids or sample-units themselves. In Charts (1) and (2) the zone-cells have been numbered serially from 1 to 96. This serial number is followed by either the letter (A) or the letter (B) to indicate the sub-sample to which the zone-cell was allotted.

72. *Field organization.* It will be convenient at this stage to explain the structure of the field organization. Each district was divided into three blocks; and each block into two sub-blocks each of which consisted of 16 zone-cells. We thus had in each district 6 sub-blocks of 16 zone-cells each. Each block was placed in charge of a chief inspector under whom there were 4 field parties. Each field party consisted of 4 investigators under one inspector. The 4 investigators in each party worked in two groups of two, called pair (1) and (2), the two pairs

---

* In Bengal these are called 'half-samples' to indicate that the complete sample was made up of two half-samples.

working in two adjoining cells. Two of the four parties in each block worked in sub-block (1) and the other two parties in sub-block (2) .Of the two parties in each sub-block, one party belonged to sub-sample (A) and the other party to sub-sample (B).

73. *Haltages.* The field survey was scheduled to start early in February 1944 and to continue for a period of 60 days which was divided into 5 haltages of 12 days each. The field work actually started on the 12th February and was closed on the 13th April, the total period covered being 62 days of which 2 days were non-working days on account of the *holi* festival.

74. *Time programme.* The time programme can be best explained with the help of the two charts (1) and (2) in which suitable code-numbers are shown in the different zone-cells. As already mentioned, out of the 16 cells in each sub-block, 8 belonging to sub-sample (A) are marked with the letter (A) and the other 8 belonging to sub-sample (B) are marked with the letter (B). The first number gives the serial number of each cell. This is followed by the letter (A) or (B) showing the sub-sample to which the cell belongs. This is followed by the number (1) or (2) showing whether the work in this zone-cell was to be done by pair (1) or pair (2). The second number (running from 1 to 5) indicates the serial number of the haltage, or the 12-day period in which any particular zone-cell was to be enumerated. Thus in Chart (1) for Monghyr at the very bottom we see the zone-cell with number 1-A-1-1 which means that the serial number of this zone-cell is (1) ; it belongs to sub-sample (A) ; and the work in this cell was to be done by party No. (1) in haltage No. (1). The adjoining cell in the same chart has the number 3-B-1-4 which indicates that the serial number of this cell is (3) ; it belongs to sub-sample (B) ; and the work in this cell was to be done by party No. (1) in haltage No. (4). The time programme was arranged in such a way as to give as far as possible a fairly even distribution of enumeration work in the different haltages over the whole district.

75. *Duplicate Enumeration.* As already mentioned, it was not possible for lack of time to adopt a fine grained sub-sampling arrangement. It was therefore considered advisable to provide other statistical controls in the design ; and it was arranged that the enumeration work would be done over again in duplicate in 25 per cent of zone-cells in the fifth haltage. In each sub-block 4 zone-cells, two belonging to sub-sample (A) and two to sub-sample (B) were selected at random and arrangements were made to have these enumerated over again by the second party in the fifth haltage. Thus the zone-cell enumerated by party (A) during the main survey would be duplicated by party (B) in the fifth haltage ; and *vice versa*. The programme of work in the fifth haltage is shown in the charts within brackets. For example, in Chart (1) in Monghyr, one cell has the number 7-A-1-2 (B-1-5). This shows that the serial number of the zone-cell is 7 ; it belongs to sub-sample (A) ; in the main survey the work here was to be done by party No. (1) in haltage No. (2) ; and also that the enumeration work in this cell would be done over again in duplicate by party (B), pair No. (1) in the 5th haltage. Further controls were provided in the form of a triplicate enumeration of from 6 to 8 per cent of the sampling-units by a number of special parties of workers. These special parties were used in the third haltage to carry out an independent enumeration of the sample-units in certain selected zone-cells in which enumeration work was done once either in the first or second haltage and in which the enumeration work was scheduled to be done again during the fifth haltage.

76. *Complete enumeration of plots.* In each block a central camp was opened and a special party of 4 investigators was stationed there for certain auxiliary work. In the first haltage of 12 days the special party in each block camp was engaged on complete enumeration of a compact area of from 50 to 60 square miles in the neighbourhood of the block camp. In this complete enumeration the investigators noted for each individual plot the proportion of land under different crops, current or old fallow etc. The primary object of collecting this information was to provide material for studying what is called the variance function by model sampling experiments in the Statistical Laboratory.

77. In the next haltage or the second 12-day period the special parties in the different block camps were interchanged so that a new party arrived at each block camp and duplicated the work done in the first 12-day period. The purpose of duplicating the complete enumeration work was two-fold. Firstly, to ensure the reliability of the primary material on which would be based subsequent studies of the variance function. Secondly, to furnish some idea regarding the degree of reliability of such primary work. (This has been discussed in Section 9).

78. In the next haltage during the third 12-day period each special party was engaged, as already mentioned, in the duplicate enumeration work in selected zone-cells. In the fourth and fifth haltages the special parties were engaged in crop-cutting work.

79. *Auxiliary parties.* Besides special parties at block camps, three auxiliary parties were used at three centres, Hilsa, Irki and Bettiah in the three districts of Patna, Gaya and Champaran respectively. Each of these parties spent 20 days at each camp in complete enumeration work. During the remaining period these auxiliary parties were engaged in crop-cutting work.

80. *Precision experiments.* Two camps for precision experiments were opened at Gogri in Monghyr district and at Bikramgunj in Shahabad district. Four technical workers (with first class M.Sc. degrees in Statistics or Mathematics) were sent from Calcutta to take charge of the work at these two centres. Harikinkar Nandi assisted by Anil Kumar Gayen was in charge of the work at Bikramgunj. Nirmal Kumar Chakravarti assisted by Gouri Sankar Chatterjee was in charge of the camp at Gogri. From five to eight investigators were posted at each of these camps; and material was collected under close supervision for studying the technique of crop-cutting work, enumeration of mixed crops and other special problems.

81. It will be noticed that the programme of work was quite comprehensive. Unfortunately, owing to the late start of the field work, the whole programme could not be completed in detail; but a great deal of valuable material was collected. The present report gives the results of the analysis of the above material as far as this has been completed; but work is still proceeding.

SECTION 4. GENERAL ACCOUNT OF THE SURVEY : RABI 1943-44

82. A general account of the crop survey in Bihar in the *rabi* season of 1943-44 is given in this Section.

STATISTICAL BRANCH

83. Shortly after my visit to Patna and personal discussions there from 16th to 18th November 1943 I was informed that the Government of Bihar had generally approved of the scheme submitted by me, and desired to have an exploratory crop survey organised in the *rabi* season of 1943-44. The statistical work was to be done by the Indian Statistical Institute under my guidance, while the field work would be done under the direct control of Government.

ARRANGEMENTS AT GIRIDIH

84. From Patna I returned to Calcutta, the headquarters of the Indian Statistical Institute, and made preliminary arrangements for starting the Bihar scheme. A portion of the Statistical Laboratory had been removed from Calcutta to Giridih as an evacuation measure early in 1942. The main library and valuable records and other equipment of the Indian Statistical Institute were housed there in premises called "Mahua" which had been handed over to the Institute free of rent for the duration of war and one year thereafter by Mrs. P. C. Mahalanobis. About 50 persons were working in the Giridih Branch at the end of November; and it was decided that the bulk of the statistical work in the Bihar scheme should be done there.

85. I went to Giridih on the 1st December accompanied by Mr. J. M. Sen Gupta to start the work on the Bihar scheme. A good deal of material in the way of Settlement Reports, Crop and Season Reports, district maps etc. required for this purpose had been collected during my visit to Patna and had been already sent to Giridih; orders were also issued by Government on the 30th November 1943 to have two sets of C. S. village maps sent there immediately.

86. *The first design for the survey.* Mr. Williams (Secretary, Revenue Department) was very keen that the eleven important wheat districts should be all included in the survey. As the total area proposed to be covered was large, something like 42,000 square miles, while the field staff would be comparatively small it was decided to adopt a two-stage design with the random selection of *mauza* maps in the first stage and location of grids at random within selected *mauza* sheets in the second stage. A good deal of labour was spent in the tabulation of crop records; preparation of maps etc.; and the first design for the survey was got ready by the first week of December 1943. C. S. village maps were, however, necessary to start the actual preparatory work of locating the grids on appropriate village maps. During my visit to Patna in November I had emphasised the importance of the C. S. village maps reaching Giridih at a very early date for this purpose. Unfortunately, there was great delay in receiving the village maps. On the 15th December maps had not yet arrived from some of the important wheat districts. It was clear that the preparatory work on the scale originally proposed could not possibly be finished in time.

87. *The second design for the survey.* During my visit to Patna on the 17th December 1943 it was, therefore, decided to give up the original scheme, and to restrict the exploratory

survey to only two districts Shahabad and Monghyr for which village maps had reached Giridih. The original design of the survey had to be abandoned; and work had to be started immediately on a new design.

88. *Lack of accommodation.* I have already mentioned that about 50 persons were working at Giridih at the end of November. A staff of at least 180 was, however, necessary to finish the preparatory work in time for the *rabi* survey. I asked Mr. N. T. Mathew (Statistician in charge of the Giridih Branch) to appoint as many local workers as he could get; and decided to transfer the required number of men from Calcutta to Giridih.

89. *Proposal for land acquisition.* Our outstanding difficulty was, however, lack of accommodation. I had pointed this out during my visit to Patna, and I had suggested that certain pieces of land adjoining the premises in the occupation of the Institute should be acquired through the Land Acquisition Act and that the Institute should put up necessary structures at its own cost. My proposal was approved in principle by the Government of Bihar, and we were assured that full Government support would be given to us in this matter. Instructions were issued accordingly to Mr. J. S. Wilcock, I.C.S., Deputy Commissioner, Hazaribagh, to help us in securing necessary accommodation for our work. He came to Giridih on the 2nd December 1943, and on the 3rd I discussed with him our immediate requirements. Pending acquisition of land, certain premises (belonging to one Mr. Myers) adjoining the Institute premises were placed at our disposal through his good offices from the 8th December 1943.

90. This made available to us some badly needed floor space but did not solve the housing problem at Giridih. We, therefore, submitted to Government on the 15th February 1944 an application for the acquisition of certain lands. During my visit to Patna at the end of March I had the opportunity of discussing this question with H. E. Mr. (now Sir) Francis Mudie, then Governor of Bihar, who very kindly assured me that immediate action would be taken in this matter by Government. A little later notification for acquisition of land for the Institute was issued, and official enquiries by the Land Acquisition Collector were held in June and July. In the meantime we purchased a large quantity of cement and iron (with permits issued by the Government of India on the recommendation of the Government of Bihar), bricks, and other building materials; but our construction programme has been held up as orders for land acquisition had not been issued by Government up to the time of writing this report. This is very seriously hampering the Bihar Crop Survey work.

PREPARATORY WORK

91. *General arrangements.* The new design for the survey (a description of which has been given in Section 3) was completed by the end of December 1943. Many details had to be attended to in organizing the preparatory work on a large scale. I went to Patna on the 7th January 1944 where I discussed matters with Rai Bahadur S. P. Sinha; we then both went to Giridih. In the meantime, local recruiting of additional hands had been proceeding at Giridih; and about 60 men were also transferred from Calcutta to Giridih. These men had not only to be housed but also fed; at one time four different kitchens had therefore to be run by the Institute for this purpose.

92. *Location of grids and preparation of field lists.* The first task to be undertaken was the sorting and grouping of the village maps by *thanas* which itself was a large undertaking. This was followed by the location at random of eighteen thousand sample-units or grids on appropriate village maps. The preparation of field lists of more than two lakhs of plots also proceeded simultaneously.

93. After marking the grids on the appropriate village maps, these were sorted into bundles among the two samples (A) or (B), six different sub-blocks, and five haltages in accordance with the design of the survey. There were also various other items of work of a routine type in getting things ready for the survey.

94. *Special difficulties.* Unforeseen difficulties also arose from time to time. One of these is worth mentioning. Maps of certain villages in Monghyr district each consisted of a very large number of separate sheets, and in one case a single village had more than 170 different sheets. No consolidated village map showing the boundaries of the different sheets was available. It was impossible to locate the grids in the correct proportion or even to identify the grids. It was necessary therefore to prepare a map of each village as a whole showing the boundary of each individual sheet.

95. This could be done in the usual way by tracing and preparing drawings to scale which however would take an enormous amount of time. A photographic method was, therefore, devised for this purpose. Each village sheet was photographed on a small scale on cinema films with a microfile camera and contact prints were made. Each print was

cut out along the actual village boundary ; and these small village contours were then fitted (something like a jig-saw puzzle) into a map of the village as a whole. This reconstructed map was then again photographed, and a print was made on an appropriate scale to fit into the *thana* map. The volume of photographic work was large as nearly 3,000 different photographs had to be made of village maps for the preparatory work in the *rabi* season alone.

96. *Completion of preparatory work.* The preparatory work had to be done at high pressure, but inspite of many difficulties it was completed in time. By the end of the first week of February 1944 all necessary maps, field lists, charts, instruction sheets etc. were ready for being despatched to the field branch according to plan.

ORGANIZATION OF THE FIELD BRANCH

97. *Preliminary arrangements.* Rai Bahadur S. P. Sinha was appointed Superintendent of Statistical Surveys just before the Christmas holidays ; but even before formally taking over charge he met me in Patna on the 17th and 18th December 1943 to discuss matters relating to the organization of the field branch. He and Mr. Priti Ranjan Sen of the Revenue Department came to Calcutta on the 25th December 1943 and stayed there for a few days to look into arrangements relating to the Crop Survey Scheme in Bengal.

98. *Loan of Bengal staff and local recruitment in Bihar.* The recruiting of the field staff was started in January 1944. Rai Bahadur Sinha suggested that as time was short it would be a distinct advantage if arrangements could be made to send on loan from Bengal a staff of experienced crop survey workers to help in giving training to the locally recruited men in Bihar and also generally in organizing the exploratory survey in the *rabi* season. I gladly agreed to this proposal, and 74 crop survey workers were sent to Bihar on loan from Bengal in charge of two field supervisors, Messrs. Dhirendra Mohan Ganguli and Pranay Kumar Chatterjee and 4 technicians, Harikinkar Nandi, Nirmal Kumar Chakravarti, Anil Kumar Gayen, and Gourishankar Chatterjee who were all first class M.SC.'s in either Statistics or Mathematics.

99. *Strength of field staff.* It will be remembered that the field programme was divided into broadly two portions, the general or extensive scheme covering the whole of the two districts Shahabad and Monghyr ; and the special (or intensive) scheme in which work was concentrated at two centres in each of the two districts Shahabad and Monghyr, where work of a more advanced nature was done. In addition, certain areas in Patna, Gaya, and Champaran districts were also covered by moving parties for work of a preliminary nature.

100. The staff employed in the different schemes is shown below in Table (4/1). Col.(1) gives the names of districts and centres; and cols.(2) to (5) the number of different categories of workers sent on loan from Bengal; col.(2) gives the number of chief inspectors; col.(3) the number of inspectors; col.(4) the number of investigators; and col.(5) the total number. Similar information is given for the staff recruited in Bihar in col.(6) to (9) respectively and for the combined staff in col.(10) to (13). The Bengal staff of 74 consisted of 6 chief inspectors, 27 inspectors and 41 investigators; the number of workers recruited in Bihar was 102 consisting of 6 chief inspectors, 22 inspectors and 74 investigators; the combined field staff in the scheme

TABLE (4/1). DISTRIBUTION OF FIELD STAFF BY DISTRICTS AND SCHEMES

Name of District	Staff on loan from Bengal				Staff recruited in Bihar				Total field staff			
	Chief inspectors	Inspectors	Investigators	Total	Chief-inspectors	Inspectors	Investigators	Total	Chief inspectors	Inspectors	Investigators	Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	GENERAL (EXTENSIVE) SCHEME											
Shahabad	3	12	12	27	3	11	24	38	6	23	36	65
Monghyr	3	12	11	26	3	11	26	40	6	23	37	66
Total	6	24	23	53	6	22	50	78	12	46	73	131
	SPECIAL (INTENSIVE) SCHEME											
Shahabad	—	1	8	9	—	—	9	9	—	1	17	18
Monghyr	—	1	8	9	—	—	10	10	—	1	18	19
Patna, Gaya, & Champaran	—	1	2	3	—	—	5	5	—	1	7	8
Total	—	3	18	21	—	—	24	24	—	3	42	45
Grand Total	6	27	41	74	6	22	74	102	12	49	115	176

was thus 176 made up of 12 chief inspectors, 49 inspectors, and 115 investigators. There were also 4 technicians (or trained statistical workers) and two field supervisors from Bengal, and the Superintendent of Statistical Surveys and his assistant in Bihar.

PROGRESS OF FIELD WORK

101. *Training of field staff.* It had been decided to give some preliminary training to the newly recruited staff in Bihar, and a training camp was opened for this purpose on the 3rd February 1944 at Gogri in Monghyr district. The six chief inspectors and a number of inspectors from Bengal in charge of two field supervisors also went there to organize the training camp, and all Bihar trainees were instructed to join there. Training was given for one week after which the field workers went to their respective centres of work. In the meantime ,the remaining workers from Bengal went to their respective camps directly from the Giridih Statistical Laboratory carrying with them all necessary maps, field lists etc.

102. *Progress of field survey.* Actual field work was started on 12th February and was continued for two months terminating on the 13th April 1944. The total period of work was 62 days out of which two days were treated as holidays on account of the *holi* festival so that the actual period of work was 60 days divided into 5 haltages of 12 days each as already noted. The field work proceeded smoothly and was completed at the scheduled times under the able supervision of Rai Bahadur S. P. Sinha. The total volume and distribution of work have been given in Section 8.

103. *Conference at Bikramganj.* I went to Patna on the 23rd March 1944 and then to Bikramganj on the 25th March and stayed there for two days for inspecting the field work. Rai Bahadur S. P. Sinha, his Assistant Supervisor, the two field supervisors from Bengal Dhirendra Mohan Ganguly and Pranay Kumar Chatterjee, had also gone there. On the statistical side Messrs. J. M. Sen Gupta (in charge of Bengal Crop Survey work), N. T. Mathew (in-charge of Bihar Crop Survey work at Giridih) and the 4 technicians, Harikinkar Nandi and Anil Gayen (in charge of the special camp at Bikramganj), Nirmal Kumar Chakravarti and Gourishankar Chatterjee (in charge of the special camp at Gogri) had also assembled at Bikramganj. We availed this opportunity to have not only a joint inspection of the field work but a series of joint discussions lasting for several hours each day on matters relating to the technique of field operations. The problem of mixed crop, and sampling procedure for crop-cutting work received special attention. Certain decisions were reached at these meetings to which reference will be made in appropriate sections.

104. *Contact between field and statistical branches.* Throughout the progress of the survey close contact was maintained between the field and statistical branches. As already mentioned, Rai Bahadur had come to Calcutta in the fourth week of December 1943; I went to Patna on the 17th January 1944; and then we both went to Giridih. Rai Bahadur came to Calcutta on the 17th February and again early in April, while I had gone to Patna and had inspected the field work at the end of March. Besides this, workers from the Statistical Laboratory went to Patna from time to time to discuss various matters.

TABULATION AND STATISTICAL ANALYSIS

105. *Original plan for tabulation of rabi records.* Instructions had been given to the field staff to send the material collected in each haltage, as soon as these were ready, directly to Giridih. At the statistical end the intention was to arrange continuous tabulation with the preparation of accumulated totals for each haltage so that up to date results of the survey would be available within a week of the receipt of the field records. As already mentioned, I went to Giridih on the 10th February and stayed there for about a week to make necessary arrangements for this purpose.

106. *Delay in arrival of field records.* The plan was however upset for two reasons. Unfortunately, there was considerable delay in the despatch of field records during the first two or three haltages owing to a very trivial hitch, namely, the lack of provision for purchasing postage stamps by the investigators. Field records began to reach Giridih in small batches from the 29th February 1944 and were tabulated as fast as possible.

107. *Change of plan.* By the third week of February certain developments had occurred which made changes in the plan inevitable. In the original scheme it had been proposed that, in case the work in the *rabi* season progressed smoothly, a survey on a somewhat larger scale would be taken up in the *bhadoi* season in which field work would start at the end of June or early July. The original plan was to complete the tabulation of the material collected in the *rabi* season by the end of April, and start the preparatory work for the *bhadoi* season sometime in early May as workers became free from *rabi* tabulation.

108. *Preparatory work for early makai (maize) survey.* Rai Bahadur S. P. Sinha had however suggested that a small scale survey of the early *makai* (maize) crop should be taken up in May in certain tracts of Monghyr, Purnea and Bhagalpur. As this would give good opportunities for strengthening the field staff I readily agreed to the proposal and Government sanction was obtained accordingly. A good portion of the Giridih staff had, therefore, to be engaged in the preparatory work for the *makai* survey.

109. *Enlarged scheme for the bhadoi survey.* Another decision made at the end of March necessitated further changes in our original plan for tabulation. The field survey in the *rabi* season had been organized so successfully that Rai Bahadur thought there would be no difficulty in attempting a survey on a full provincial scale in the *bhadoi* season of 1944 instead of waiting till the *aghani* season of 1945 as proposed in the original scheme. Examination of the *rabi* records showed that quality of field work had been quite satisfactory; and personal inspection of the field work at the end of March convinced me of the soundness of the Rai Bahadur's proposal. Government sanction was, therefore, obtained and necessary arrangements were made for starting the preparatory work for the survey in the *bhadoi* season from the begining of April.

110. *Slow tabulation of rabi records.* Thus, owing, first, to the initial delay in receiving the field records; and secondly, the starting of preparatory work for early *makai* and *bhadoi* surveys, the staff engaged at Giridih on the *rabi* material had to be reduced appreciably. Further increase of staff at Giridih was not possible for lack of accommodation. A portion of the Bihar work was therefore transferred to Calcutta which caused delays. I went to Giridih on the 12th April and stayed there for three days to see to local arrangements. As priority had to be given to the preparatory work for *makai* and *bhadoi* surveys, slowing down of the tabulaion of the *rabi* materials became inevitable.

111. *Preliminary report.* A preliminary report on the survey in the *rabi* season based on material collected in three (out of five) haltages, that is, up to 28th March was, however, submitted on the 28th April 1944.

112. *Subsequent delays.* It was my intention to submit another report in the course of a month but this was not possible for unforeseen reasons. The primary tabulation of *rabi* had been completed by the middle of May, and I started writing the report at Kalimpong on the 16th May ,1944. I was, however, obliged to go down suddenly to Calcutta on the 19th May to discuss matters relating to the Bengal Crop Survey Scheme with the Government of Bengal. Unfortunately, I had an acute attack of appendicitis at the end of May which interrupted work on the present report. I was also obliged to cancel my visit to Patna for an interview which H. E. Sir T. G. Rutherford ,Governor of Bihar, had kindly agreed to give me on the 1st June. Continued ill health as well as pressing work in connexion with the Bengal Crop Survey Scheme (which had been sanctioned in the meantime on a much enlarged scale for three years) interfered with the writing of the present report. Another difficulty was that in Calcutta I was unable to examine the primary records and tabulation sheets which were at Giridih; while, owing to my illness, I was unable to go to Giridih for several months.

113. *The present report.* I arrived at Giridih on the 2nd October and immediately started writing this report; but there was another interruption as I had to go to Calcutta in the 17th October and stay there for about a week for some urgent statistical work of the Government of Bengal. The report was, however, finished on the 31st October, much later than I had originally intended.

ACKNOWLEDGEMENTS

114. Without the ungrudging help of my fellow-workers in the Indian Statistical Institute a scheme of this kind could not have been successfully carried out. Mr. J. M. Sengupta helped me all along in the planning and the superior direction of my work; the brunt of the work fell on Mr. N. T. Mathew who was in charge of the statistical work at Giridih, and was ably assisted by the local staff, among whom Jitendranath Taluqdar, Haribhajan Choudhury, Rajendrachandra Roy, Dhirendrachandra Sarker, and Probhat Ranjan Sinha deserve special mention. Mr. Brajokishore Sinha was responsible for finishing in time the heavy photographic work at the preparatory stages. In the field branch good work was done by the two supervisors, Messrs. Dhirendra Mohan Ganguly and Pranay Kumar Chatterjee and the four technicians, Harikinkar Nandi, Nirmal Kumar Chakravarti, Anil Kumar Gayen and Gouri Shankar Chatterjee whose services were sent on loan to Bihar from the Institute.

PART 2. RESULTS OF THE RABI SURVEY

SECTION 5. AREA UNDER RABI CROPS : 1943-44

115. The present section gives the results relating to the area survey which, as already noted, began rather late in the season. The programme of work was divided into five haltages each covering a 12-day period. In certain regions harvesting started from the third haltage (between 8th and 20th March 1944) or practically from the middle of the area survey programme. Area survey along with crop-cutting work was continued during the fourth (21st March -2nd April) and fifth (3rd April -13th April) haltages, but in the latter haltage a good deal of the crop was found to have been already harvested. This naturally affected the results of the area survey; records for *masuri* were mostly missing as much of the crop had been harvested before the area enumeration could begin; and this crop has been omitted from the present report.

CROP ACREAGE BY HALTAGES

116. Table (5/1) gives the area (in thousand acres) of five important crops (wheat, barley, gram, *arhar*, and *khesari*) by districts and haltages. The serial number of the haltages is shown in col.(1); the period (date) of survey in col.(2); and the total number of grids or sample-units enumerated in col.(3). The area in thousand acres is given in successive cols. (4) to (8) for the five crops, wheat, barley, gram, *arhar*, and *khesari* respectively.

TABLE (5/1). CROP AREA IN THOUSAND ACRES BY DISTRICTS AND HALTAGES

Serial No. of haltages	Period of Enumeration	Number of grids	Area in thousand acres under				
			Wheat	Barley	Gram	Arhar	Khesari
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DISTRICT : SHAHABAD (AREA = 2,798,000 ACRES)							
1	11 Feb - 23 Feb	1607	261 ·95	98 ·31	429 ·65	81 ·65	276 ·93
2	25 Feb - 6 Mar	1636	360 ·45	81 ·11	356 ·48	63 ·72	273 ·77
3	8 Mar - 20 Mar	1596	303 ·43	71 ·11	290 ·50	53 ·55	138 ·58
4	21 Mar - 2 Apl	1433	276 ·47	56 ·13	241 ·51	35 ·41	101 ·68
5	3 Apl - 13 Apl	1051	135 ·16	18 ·47	89 ·55	57 ·09	24 ·35
1-5	11 Feb - 13 Apl	7323	276 ·64	69 ·80	301 ·77	58 ·38	172 ·91
DISTRICT : MONGHYR (AREA = 2,370,000 ACRES)							
1	12 Feb - 25 Feb	1408	307 ·22	65 ·49	292 ·81	144 ·30	98 ·91
2	26 Feb - 7 Mar	1539	244 ·94	71 ·70	190 ·05	109 ·56	22 ·07
3	8 Mar - 20 Mar	1413	255 ·71	65 ·42	141 ·37	77 ·64	28 ·17
4	21 Mar - 2 Apl	1488	327 ·57	65 ·03	149 ·43	83 ·37	42 ·07
5	3 Apl - 13 Apl	1181	188 ·16	24 ·41	123 ·23	73 ·70	19 ·20
1-5	12 Feb - 13 Apl	7029	267 ·96	59 ·77	181 ·64	98 ·67	43 ·00

117. The chief point to be noted in this table is the steadily decreasing area under practically all the different crops in successive haltages. This was presumably due to the increasing proportion of the crop having been harvested in the later haltages. The lesson to be learnt from the present records is that the area survey in the *rabi* season must be completed before the beginning of harvesting which in practice would mean by the end of February or early March. It would, however, be obviously desirable to carry out a skeleton area survey at the time of harvesting to determine what proportion of the total area sown is actually harvested. This point has been further discussed in a later section (paragraphs 318, 319).

CROP ACREAGE BY DISTRICTS AND SUB-SAMPLES

118. For reasons explained above, the records for the first four haltages only have been taken into consideration for estimating the district crop acreage figures and their margin of error which are given in Table (5/2). Col. (1) gives the name of the crop; col. (2) the area (in thousand acres) under each crop (together with the corresponding standard error) estimated from sample (A); col. (3) gives the corresponding estimate based on sample (B). The difference between the two estimates (A) and (B) expressed in terms of the standard error of the difference is given in col. (4) for convenience of statistical comparison. Combined estimates of crop acreage based on both samples (A) and (B) are shown in col. (5).

Corresponding acreage figures as given in the Settlement Reports (covering the years 1907-1916 in Shahabad and 1905-1912 in Monghyr) are shown in col. (6); and corresponding official forecasts for 1943-44 (taken from the fourth wheat forecast and that dated 26th April 1944 for other crops issued by the Director of Agriculture) are shown in col.(7).

TABLE (5/2). CROP AREA IN THOUSAND ACRES BY DISTRICTS AND SUB-SAMPLES (BASED ON FIRST FOUR HALTAGES).

Crop	Sample(A)	Sample(B)	d(A—B)	Combined(A+B)	Settlement Report (a)	Forecast 1943-44 (b)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
	DISTTICT : SHAHABAD (AREA = 2,798,000 ACRES)					
	N = 3307	N = 2965		N = 6272		
Wheat	296·36±27·13	304·37±29·16	0·20	300·35±19·87	227·9	263·80
Barley	74·63±16·10	82·08± 9·23	0·40	78·40± 9·51	105·4	102·50
Gram	369·02±42·75	305·55±33·07	1·17	337·33±26·86	339·5	282·30
Arhar	62·06±16·51	55·14±18·45	0·28	58·60±12·59	51·4	41·00
Khesari	251·85±33·18	148·75±34·25	1·56	197·81±23·78	—	—
	DISTRICT : MONGHYR (AREA = 2,370,000 ACRES)					
	N = 3165	N = 2683		N = 5848		
Wheat	242·25±35·23	325·99±40·22	1·57	284·08±31·62	144·9	219·80
Barley	44·86± 5·80	88·94± 7·11	4·80	66·90± 5·32	81·4	81·90
Gram	222·24±36·87	164·27±26·39	1·27	193·43±26·86	210·5	210·50
Arhar	105·17±25·52	102·21±29·90	0·08	103·71±23·22	(c)42·5	81·00
Khesari	45·81±16·75	49·82±19·69	0·16	47·80±15·11	—	—

(a) Settlement period : Shahabad 1907—1916 and Monghyr 1905—1912.

(b) Forecast of *rabi* crops dated 26th April 1944, and Fourth wheat forecast issued by Director of Agriculture.

(c) *Arhar* figure for North Monghyr alone in Settlement Report.

119. In Shahabad district the sample survey estimates of crop acreage (in thousand acres)are 300·35±19·87 for wheat, 78·40±9·51 for barley, 337·33±26·86 for gram, 58·60±12·59 for *arhar*, and 197·81±23·78 for *khesari*. In Monghyr district the corresponding values were 284·08±31·62 for wheat, 66·90±5·32 for barley, 193·43±26·86 for gram, 103·71±23·22 for *arhar* and 47·80±15·11 for *khesari*. We have already seen, however, that appreciable portions of the crop had been harvested during the later stages of the area enumeration; above figures must therefore be to some extent under-estimates; and are thus not comparable with official figures.

OFFICIAL FIGURES NOT STRICTLY COMPARABLE

120. There is another serious difficulty in comparing sample and official estimates. In Bihar in the *rabi* season most of the plots are sown with mixed crops; but it is not at all clear what is exactly meant by the area under any individual crop sown in mixture with other crops. The subject is full of complications and has been discussed in a separate section.

121. For reasons explained above a valid comparison is not possible between sample and official figures. A very rough comparison can, however, be made for just what it may be worth. In Shahabad district with a total area of 2,798,000 acres we find that sample survey gives an estimate of about 300,000 acres under wheat against a Settlement figure of about 228,000 and an official forecast of 263,800 acres in 1943-44. The Settlement figures are very old and refer to the period 1907-16 for Shahabad and 1905-12 for Monghyr. Considerable changes in crop acreage may have easily taken place since then. It is, therefore, quite possible that the area under wheat had actually increased to about 300,000 acres in 1943-44, so that the official

forecast of 263,800 acres may be appreciably in defect. For barley the area during 1907-1916 was 105,420 acres according to the Settlement Report; the official estimate gives practically the same figure while the sample estimate of 78,400 acres is somewhat lower, but nothing improbable. For gram, the sample estimate of 337,330 acres is very nearly the same as the Settlement figures of 339,510 acres; the official forecast of 282,300 acres is appreciably lower. For *arhar* also the sample figure of 58,600 acres is nearer the Settlement figure of 51,440 acres than the official forecast of 41,000 acres. Neither Settlement nor official figures are available for *khesari*.

122. In Monghyr the sample estimate of the wheat area of 284,080 acres is appreciably higher than both the Settlement figure of 144,860 acres and the official forecast of 219,800; this is quite plausible. Barley with a sample estimate of 66,900 acres against Settlement figures of 81,430 acres and official forecast of 81,900 acres, shows a shrinkage as in Shahabad district. For gram the sample estimate of about 193,000 acres is somewhat smaller than the two official figures of about 211,000 acres. But for *arhar* the sample estimate of 103,710 acres is much larger and of quite a different order from the official forecast of only 1,000 acres .It is interesting to note that according to the Settlement Report(1905-07, Appendix D(ii)) the acreage under *arhar* in North Monghyr alone was about 42,000 acres so that the sample estimate for the whole district is not at all improbable.

123. On the whole, from the sample figures it looks likely that the area under wheat has increased appreciably and is greater at present than at the time of Settlement operations; area under barley has probably decreased somewhat; area under gram has remained about the same as in previous years; and area under *arhar* has increased in both districts. The official forecast of 1,000 acres for *arhar* in Monghyr is entirely wrong. Settlement figures are believed to be generally reliable, and may therefore be accepted as respresenting actual crop acreages at the time of Settlement operations. Since then changes must have taken place; but against the background of Settlement figures the sample estimates look quite plausible indicating that they have an objective foundation.

AGREEMENT BETWEEN SAMPLE ESTIMATES

124. This is all that can be said so far as the external evidence is concerned. We may now consider the internal evidence in favour of the sample survey. I have already explained in an earlier section how the whole survey was organized in two independent portions. In each district, the total number of grids or sample-units was divided into two different portions sample (A) and sample (B)—scattered over the whole district. In Shahabad out of a total number of 6,272 grids there were 3,307 grids in sample (A) and 2,965 grids in sample (B). In Monghyr district out of 5,848 grids there were 3,165 in sample (A) and 2,683 in sample (B). The grids in each of the two samples (A) and (B) were scattered over the whole district and were enumerated independently by two different sets of investigators. Each of the two samples (A) and (B) can thus furnish an independent estimate of the crop acreage for the district as a whole.

125. Let us consider actual figures. In Shahabad district according to sample (A) the area under wheat (in thousand acres) was 296·36±27·13 as shown in col. (2) of Table (5/2). (The latter figure of 27·13 is the expected sampling error of the actual estimate of 296·36). According to sample (B) the area under wheat (in thousand acres) in Shahabad district was 304·37±29·16 (the latter figure 29·16 again being the estimated margin of error of 304·37). These two estimates 296·36 and 304·37 are, of course, not identical. They are each based not on complete enumeration of all plots under wheat but on only a comparatively small fraction of such plots. They are therefore both subject to fluctuations due to sampling. In fact the estimated standard errors of ±27·13 and ±29·16 (thousand acres) for the two samples (A) and (B) respectively are statistical measures of the magnitude of such fluctuations due to sampling.

126. The difference between the two estimates 296·36 and 304·37 is 8·01 (thousand acres). The estimated margin of error in the case of either sample is, however, appreciably greater than the observed difference. We find therefore that the two sample estimates (A) and (B) for wheat acreage in Shahabad district are in agreement within the margin of error, that is, within the limits of errors of sampling. Such internal agreement must be considered to be good evidence of the reliability of the sample estimates.

127. Without entering into technical details it may be stated here that for a statistical comparison of the two sample estimates (A) and (B) it is convenient to express the difference between the two estimates in terms of the margin of error of the difference itself. These values (called 'd' for convenience of reference) have been shown in col. (4) of Table (5/2). It will be noticed that the value of 'd' is less than 1 in five cases and lies between 1 and 2 in

four cases ; only in one single case, namely, barley in Monghyr district, the value of 'd' is comparatively large. Thus, with one single exception, the two estimates based on samples (A) and (B) are in satisfactory statistical agreement.

128. For barley in Monghyr district the two estimates are 44·86±5·80 and 88·94±7·11 (thousand acres) for samples (A) and (B) respectively. The difference is admittedly large and a value of 'd' of 4·80 cannot be neglected. It is necessary however to recall at this stage that the division into the two samples (A) and (B) was made on the basis of comparatively large zone-cells (each of area 50 square miles approximately), and was not carried to individual sample-units. As already explained in an earlier section, this rough and course grained structure of the two interpenetrating sub-samples makes it possible for real geographical differences in the different zone-cells having been shown up in the two sample estimates. That is, the two estimates (A) and (B), although mixed up for the district as a whole, refer to fairly big patches in geographically distinct areas. The observed difference for barley in the two estimates (A) and (B) may therefore be due to real regional differences in the intensity of cultivation in the big patches into which the samples were divided. It is thus not unreasonable to conclude that the two sets of results based on the two samples (A) and (B) are in satisfactory agreement and hence the sample estimates are on the whole reliable (within, of course, respective limits of sampling errors.)

129. In view of the importance of the question, we may also consider the agreement between the two samples (A) and (B) in closer detail separately for each sub-block in each district. Relevant material is given in Table (5/3) in which col. (1) gives the serial number of the sub-block ; cols. (2) and (3) give for Shahabad district the number of grids or sample-units in samples (A) and (B) respectively ; cols. (4) and (5) the crop acreage with corresponding standard errors based on samples (A) and (B) respectively ; and col. (6) the value of 'd', that is, the difference between the two estimates (A) and (B) divided by the standard error of the difference. Similar data for Monghyr district are given in cols. (7) to (11).

130. It will be noticed that for individual sub-blocks the agreement between estimates based on samples (A) and (B) are generally satisfactory as judged by the values of 'd'. Out of a total number of 60 values of 'd' given in cols. (6) and (11) we notice that 51 are less than 1, and the remaining 9 lie between 1 and 2. We also find that 5 out of 6 values of 'd' are greater than 1 for barley in Monghyr district ; the differences in the sub-blocks are not unduly large ; but sample (B) is usually giving higher estimates with the result that the difference for the district as a whole is quite large indicating the possible existence of real differences in the intensity of cultivation of barley between different zone-cells of Monghyr.

ANALYSIS OF VARIANCE FOR AREA SURVEY

131. Certain technical data relating to the precision of the results of the area survey are given in Table (5/4) in which the name of the crop is shown in col. (1). This is followed by figures relating to sample (A) in cols. (2) to (5), and corresponding data relating to sample (B) in cols. (6) to (9), and finally for the two samples combined together in cols. (10) to (13). Cols. (2), (6) and (10) give the mean proportion of land under each crop ; cols. (3), (7) and (11) the variance between (mean proportion of) zone-cells ; and cols. (4), (8) and (12) the variance within zone-cells. Finally the ratio of the variance between zone-cells to the variance within zone-cells (V_1/V_2) is shown for sample (A) in col. (5), for sample (B) in col.(9), and finally for the two samples (A) and (B) combined together in col. (13). It is worth noting that the ratio of the two variances given in cols. (5), (9) and (13) is on the whole large showing that the greater part of the fluctuation in the intensity of cultivation occurs from one zone-cell to another.

132. Based on the data given in Table (5/4) the standard deviations for different stages calculated by the appropriate two-stage formula are given in Table (5/5). The name of the crop is shown in col. (1). This is followed by the mean proportion of land under each crop for sample (A) in col. (2) ; calculated standard deviation of fluctuations between cell-means (S_1) in col. (3), and the standard deviation of fluctuations within zone-cells (S_2) in col. (4). Corresponding figures for sample (B) are given in cols. (5), (6) and (7) respectively. The mean proportion of land under cultivation (with standard error) for each crop for the two samples (A) and (B) combined together is given in col. (8) ; and the percentage variability of the combined proportion in col. (9). It will be noticed that the over-all margin of error for the estimates of crop acreages for individual districts lie between roughly 7 and 12 per cent for wheat, barley and gram ; but are appreciably higher for *arhar* and *khesari*.

Table (5/3 Comparison of crop area (in thousand acres) by Sub-samples and sub-blocks

Serial No. of sub-blcok	District Shahabad: Number of grids (A)	District Shahabad: Number of grids (B)	District Shahabad: Estimated area with standard error Sample (A)	District Shahabad: Estimated area with standard error Sample (B)	Diff. d(A-B)	District Monghyr: Number of grids (A)	District Monghyr: Number of grids (B)	District Monghyr: Estimated area with standard error Sample (A)	District Monghyr: Estimated area with standard error Sample (B)	Diff. d(A-B)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					Wheat					
1	513	528	62·90±16·48	68·82±15·74	−0·26	517	632	1·50±19·56	6·53±17·08	−0·19
2	579	489	61·84±14·21	62·82±15·82	−0·05	614	511	13·68±16·13	17·74±17·92	−0·17
3	594	552	40·54±13·56	60·83±14·50	−1·03	654	569	57·76±18·02	46·95±19·68	+0·41
4	677	379	36·47±15·23	33·21±22·85	+0·12	508	426	59·56±20·03	85·84±22·53	−0·87
5	354	540	69·85±19·60	65·57±14·85	+0·17	423	234	44·68±21·81	76·24±28·43	−0·88
6	590	477	24·76±16·37	13·12±19·22	+0·46	449	311	65·07±21·90	92·69±25·18	−0·83
	3307	2965	296·36±27·13	304·37±29·16	−0·20	3165	2683	242·25±35·23	325·99±40·22	−1·57
					Barley					
1	513	528	19·28± 5·12	23·74± 4·94	−0·63	517	632	1·65± 4·33	9·13± 3·86	−1·29
2	579	489	17·20± 4·46	15·41± 4·90	+0·27	614	511	5·10± 3·59	12·07± 4·02	−1·29
3	594	552	9·14± 4·25	7·21± 4·50	+0·31	654	569	3·10± 4·07	2·59± 4·41	+0·09
4	677	379	4·66± 4·77	9·22± 7·10	−0·53	508	426	9·19± 4·51	16·52± 5·04	−1·08
5	354	540	15·71± 6·09	13·08± 4·62	+0·34	423	234	3·20± 4·88	15·06± 6·40	−1·47
6	590	477	8·64± 5·08	13·42± 6·00	−0·61	449	311	22·62± 4·88	33·57± 5·64	−1·46
	3307	2965	74·63±16·10	82·08± 9·23	−0·40	3165	2683	44·86± 5·80	88·94± 7·11	−4·80
					Gram					
1	513	528	81·53±14·29	86·23±13·64	−0·24	517	632	1·46±15·82	9·05±13·86	−0·37
2	579	489	82·06±12·34	52·97±13·72	+1·58	614	511	12·14±13·04	22·26±14·51	−0·59
3	594	552	63·23±11·76	59·80±12·57	+0·20	654	569	80·75±14·59	64·38±15·91	+0·76
4	677	379	54·09±13·21	42·59±19·77	+0·48	508	426	75·72±16·20	28·33±17·77	−1·96
5	354	540	54·14±16·97	44·03±12·86	+0·47	423	234	32·07±17·64	34·73±23·02	−0·02
6	590	477	33·97±14·19	19·93±16·73	+0·66	449	311	20·10±17·71	5·52±20·34	+0·54
	3307	2965	369·02±42·75	305·55±33·07	+1·17	3165	2683	222·24±36·87	164·27±26·39	+1·27
					Arhar					
1	513	528	32·78± 9·50	24·40± 9·03	+0·64	517	632	0·16±15·55	10·86±13·62	−0·52
2	579	489	18·27± 8·20	15·86± 9·09	+0·20	614	511	2·16±12·82	3·41±14·20	−0·07
3	594	552	3·17± 7·81	6·18± 8·32	−0·26	654	569	15·48±14·34	10·69±15·65	+0·22
4	677	379	3·16± 8·76	0·83±13·16	+0·15	508	426	1·73±15·92	9·47±17·93	−0·32
5	354	540	0·26±11·27	2·07±10·02	−0·12	423	234	56·66±17·36	57·63±22·63	−0·03
6	590	477	4·42± 9·41	5·80±11·03	−0·10	449	311	28·98±17·43	10·15±21·98	+0·67
	3307	2965	62·06±16·51	55·14±18·45	+0·28	3165	2683	105·17±25·52	102·21±29·90	+0·08
					Khesari					
1	513	528	26·21±19·65	16·02±18·76	+0·38	517	632	0·00	0·00	0·00
2	579	489	56·76±16·97	47·44±78·84	+0·37	614	511	1·18± 8·44	11·21± 9·38	−0·79
3	594	552	40·41±16·13	41·96±17·29	−0·07	654	569	22·60± 9·42	22·52±10·31	+0·01
4	677	379	25·75±18·13	8·29±27·25	+0·53	508	426	9·15±10·48	3·71±11·81	+0·34
5	354	540	84·31±23·40	29·18±17·70	+1·88	423	234	1·29±11·43	0·47±14·91	+0·04
6	590	477	18·41±19·48	5·86±22·93	+0·58	449	311	11·59±11·47	11·91±13·15	−0·02
	3307	2965	251·85±33·18	148·75±34·25	+0·28	3165	2683	45·81±16·75	49·82±19·69	−0·16

TABLE (5/4). PROPORTION OF LAND UNDER DIFFERENT CROPS BY SUB-SAMPLES : ANALYSIS OF VARIANCE BETWEEN AND WITHIN ZONE-CELLS.

Name of crops	Sample (A)				Sample (B)				Combined (A & B)			
	Mean proportion (p)	Between cells (v_1)	Within cells (v_2)	Ratio $\left(\frac{v_1}{v_2}\right)$	Mean proportion (p)	Between cells (v_1)	Within cells (v_2)	Ratio $\left(\frac{v_1}{v_2}\right)$	Mean proportion (p)	Between cells (v_1)	Within cells (v_2)	Ratio $\left(\frac{v_1}{v_2}\right)$
(1)	(2	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
DISTRICT : SHAHABAD (NUMBER OF GRIDS = 6272)												
Wheat	·1059	·2904	·0228	12·73	·1088	·2980	·0276	10·72	·1073	·3005	·0251	11·97
Barley	·0267	·1019	·0044	23·16	·0293	·0319	·0051	6·25	·0280	0·663	·0048	13·81
Gram	·1319	·7190	·0294	24·46	·1092	·3803	·0243	15·65	·1206	·5514	·0270	20·42
Arhar	·0222	·1068	·0047	22·72	·0197	·1192	·0089	13·39	·0209	·1122	·0079	14·20
Khesa	·0900	·4332	·0249	17·40	·0514	·4034	·0176	22·92	·0707	·4207	·0214	19·66
DISTRICT : MONGHYR (NUMBER OF GRIDS = 5848)												
Wheat	·1022	·6431	·0248	25·93	·1375	·6691	·0308	21·72	·1199	·6527	·0276	23·65
Barley	·0189	·0183	·0042	4·36	·0375	·0220	·0077	2·86	·0282	·0230	·0058	3·97
Gram	·0938	·7034	·0212	33·18	·0695	·2875	·0149	19·30	·0816	·4994	·0183	27·29
Arhar	·0444	·3367	·0149	22·60	·0431	·3702	·0146	25·36	·0438	·3521	·0148	23·79
Khesari	·0193	·1479	·0072	20·54	·0210	·1608	·0072	22·33	·0202	·1528	·0072	21·22

TABLE (5/5). MULTI-STAGE ESTIMATES OF THE STANDARD DEVIATION OF MEAN PROPORTION OF LAND UNDER CROPS BY SUB-SAMPLES.

Name of crops	Sample (A)			Sample (B)			Combined (A+B)	
	Mean proportion (p)	Between cells (s_1)	Within cells (s_2)	Mean proportion (p)	Between cells (s_1)	Within cells (s_2)	Mean proportion with standard error (p)	Percentage variability
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
DISTRICT : SHAHABAD (NUMBER OF GRIDS = 6272)								
Wheat	·1059	·0628	·1500	·1088	·0660	·1661	·1073 ± ·0071	6·62
Barley	·0267	·0380	·0663	·0293	·0200	·0714	·0280 ± ·0034	12·14
Gram	·1319	·1000	·1717	·1092	·0762	·1559	·1206 ± ·0096	7·96
Arhar	·0222	·0380	·0887	·0197	·0425	·0943	·0209 ± ·0045	21·53
Khesari	·0900	·0768	·1716	·0514	·0795	·1327	·0707 ± ·0085	12·02
DISTRICT : MONGHYR (NUMBER OF GRIDS = 5848)								
Wheat	·1022	·0970	·1575	·1375	·1073	·1755	·1199 ± ·0113	9·42
Barley	·0189	·0148	·0648	·0375	·0167	·0877	·0282 ± ·0019	6·74
Gram	·0938	·1020	·1476	·0695	·0700	·1221	·0816 ± ·0096	11·76
Arhar	·0444	·0700	·1221	·0431	·0800	·1208	·0438 ± ·0083	18·95
Khesari	·0193	·0061	·0848	·0210	·0053	·0848	·0202 ± ·0054	26·73

DECREASE IN CROP AREA ENUMERATED IN SUCCESSIVE HALTAGES

133. The gradual decrease in the area under practically all the crops in successive haltages has been pointed out in an earlier paragraph. It was also noted that this was probably due to the fact that harvesting had commenced in the middle of the area survey so that much of the crop had already been harvested before the area enumeration work could start in the fourth and fifth haltages. A second factor also might have been in operation. Owing to failure of rain, damage caused by pests or animals, and other accidental causes, portions of the area sown go out of cultivation and are not harvested. The area harvested in fact must be always less than the area sown. This gap between the area sown and the area harvested may be large for paddy according to Sir John Hubback and has been further discussed in a later section. I have no knowledge, and I have not been able to secure any reliable information as to the extent to which the harvested area falls short of the area sown in the *rabi* season in Bihar. If this gap is considerable then the observed decrease in acreage-estimates the later haltages may have also to be partly ascribed to this factor.

134. Although it is not possible from the available material to disentangle how far the observed decrease in area estimates is due (*a*) to previous harvesting of the crop or (*b*) to fields having gone out of cultivation through the operation of natural causes, it is of some interest to consider the facts as observed. Relevant material is given in Table (5/6) in which col. (1) gives the serial number of the haltage ; col. (2) the actual period of survey. The observed crop acreage (in thousand acres) is given in col. (3) for wheat and in cols. (5), (7), (9) and (11) for barley, gram, *arhar* and *khesari* respectively.

Table (5/6). Estimated crop area in thousand acres by haltages

Serial No. of haltage	Period of survey	Wheat		Barley		Gram		Arhar		Khesari	
		observed	graduated	observed	graduated	observed	graduated	observed	graduated	observed	graduated
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	District : Shahabad										
1	11 Feb - 23 Feb	262	335	98	102	430	452	82	73	277	298
2	25 Feb - 6 Mar	360	301	81	83	386	370	64	65	274	230
3	8 Mar - 20 Mar	303	267	71	65	290	288	54	58	139	163
4	21 Mar - 2 Apl	276	233	56	46	242	205	35	50	102	95
5	3 Apl - 13 Apl	135	200	18	23	90	123	58	43	24	27
	District : Monghyr										
1	12 Feb - 25 Feb	307	296	65	76	293	255	144	132	99	70
2	26 Feb - 7 Mar	245	280	72	67	190	217	110	115	22	56
3	8 Mar - 20 Mar	256	265	65	58	141	179	78	98	28	42
4	21 Mar - 2 Apl	328	249	65	49	149	141	83	81	42	28
5	3 Apl - 13 Apl	188	234	60	40	182	103	99	64	43	14

135. The observed acreages were graduated by simple straight lines, and the graduated values are given in cols. (4), (6), (8), (10) and (12) of the same table for wheat, barley, gram, *arhar* and *khesari* respectively. The graduating equation in each case was of the form : y=a+b(x), where (*y*) is the graduated value of crop acreage and (*x*) is the serial number of the haltage. The values of (*a*) and (*b*) for the different crops are shown in Table (5/7) separately for the two districts Shahabad and Monghyr.

Table (5/7). Linear graduation of crop area : y=a+b(x)

y = area under crop in thousand acres
x = serial number of haltages

Name of crop	Estimated values of parameters			
	'a'		'b'	
	Shahabad	Monghyr	Shahabad	Monghyr
(1)	(2)	(3)	(4)	(5)
Wheat	368 ·60	311 ·30	−33 ·80	−15 ·50
Barley	120 ·30	48 ·90	−18 ·50	−8 ·90
Gram	534 ·80	293 ·30	−82 ·40	−38 ·10
Arhar	80 ·60	148 ·50	− 7 ·60	−16 ·90
Khesari	360 ·00	84 ·00	−67 ·80	−14 ·00

SAMPLE FIGURES PROBABLY UNDER-ESTIMATES

136. On account of the progressive decrease in crop acreage in successive haltages it is clear that the average estimate based even on the first four haltages must necessarily be under-estimates of the total area sown. If the decrease in crop acreage is mainly due to the crop having been harvested before the area enumeration had taken place (which is likely) then the average sample estimates considered here must also be under-estimates of the area harvested. One could perhaps use the graduated values for the first haltage as adjusted crop acreages free from the decreasing tendency noticed in the observed values. On this

basis, the area under wheat would be 335,000 acres against an average observed value of 300,000 acres in Shahabad ; and 296,000 acres against observed average of 284,000 acres in Monghyr district. For barley we get adjusted values of 102,000 and 76,000 acres against average values of 78,000 and 67,000 acres; for gram the adjusted values of 452,000 and 255,000 acres against average values of 337,000 and 193,000 acres ; for *arhar* adjusted values of 73,000 and 132,000 acres against average values of 59,000 and 104,000 ; and finally adjusted values of 298,000 and 70,000 acres for *khesari* against average values of 198,000 and 48,000 acres in the two districts Shahabad and Monghyr respectively. These adjusted values may be accepted as upper limits of 'area sown' ; but much reliance cannot be placed on these results in the absence of factual knowledge regarding shrinkage of area sown with the progress of the crop season.

SECTION 6. RATE OF YIELD AND OUTTURN OF WHEAT AND GRAM : 1943-44

137. Crop-cutting work started in some of the sub-blocks in the third haltage, that is, between 8th and 20th March and was carried on in full swing during the last two haltages, that is, from 21st March to 13th April, 1944.

138. *Work on two crops, wheat and gram.* During preliminary discussions with the Government of Bihar I had enquired whether at least portions of the crop-cutting work could be done by (or under the supervision of) the staff of the Agriculture or Co-operative Departments, but I was informed that this was not possible. The whole of the work had to be done within the limited resources (in men and money) of the present scheme. It was decided, therefore, to restrict the crop-cutting work to only two crops, namely, wheat and gram. Even with two crops the total volume of work which could be done was strictly limited. There were only two alternatives, either to reduce the total number of cuts or to reduce the size (area) of each cut. The total number of cuts could not, however, be reduced too much as the margin of error of the average yield for the district would be too high.

139. *Size of cuts.* Extensive crop-cutting work in Bengal had shown that sample-units of a comparatively small size were more efficient in the sense that a higher precision could be attained in the final results with less expenditure of labour (and hence at lower cost). The advantage of using comparatively small cuts had also been fully established by a classical series of crop-cutting experiment conducted in Bihar by Sir John Hubback between 1923 and 1925. He had used small triangular cuts of size 1/3200 acre (=13.6 sq. feet) each ; and in his paper on "Sampling for Rice Yield in Bihar and Orissa" (Government of India, 1927) he had very clearly stated (p. 9) that there was no need of taking large samples instead of the handy samples used by him ; and had explained that this was due to the fact that, the existence of high correlation in the yield of contiguous pieces of land, "made the ordinary rule that the standard deviation of the mean was the standard deviation of the population divided by the square root of the numbers of samples quite inapplicable".

140. It is true that Sir John Hubback's work was done on paddy ; but his arguments are equally valid for other crops. In fact, the existence of high correlation between the yield of contiguous portions of land has been conclusively established for wheat and other crops by the work of Fairfield Smith (*Jour. of Agri. Sc.* Vol. 28, 1938, pp. 1-23) and others. Extensive crop-cutting work on wheat in England had also directly shown that cuts of a comparatively small size were convenient and suitable. For example, F. Yates in his articles on "Crop estimating and forecasting : Indications of the sampling observations on wheat" in the *Journal of the Ministry of Agriculture of the United Kingdom* (Vol. 43, 1936-37, p. 161) stated :—

> "......In order to provide objective estimates of yields of commercial fields ......a very simple sampling technique was adopted, only about 5 square yds. in all of any field being sampled. This amount of sampling proved amply sufficient when the purpose of sampling is that of estimating the mean yield of a district, sampling errors being small in relation to the variability between fields."

141. For reasons explained above it was decided to work with sample-cuts of 25 sq. feet each and to take two such cuts from each plot included in the sample. (This gave a sampled area of 50 square feet against a sampled area of 5 sq. yds or 45 sq. feet in the United Kingdom in the work described by Yates). The plots (from which the crop was to be cut) were selected beforehand at random in the Statistical Laboratory from among the plots which were included in the grids or sample-units used in the area enumeration. Within each selected grid two plots, one of gram and one of wheat, were picked up at random for this purpose and within each selected plot two cuts (called a doublet) were actually harvested by the worker in the way explained below.

142. *Location of cuts.* The workers were supplied with tables of random numbers out of which two numbers were picked up in serial order for each selected plot. The first random number gave the number of steps the worker was to measure along the length of the plot from one corner, and the second random number gave the number of steps which he was required to walk into the plot perpendicular to the length of the plot and parallel to the other sides. The point located in such a way formed one of the vertices of the triangular apparatus described below.

143. *Crop-cutting procedure.* Instead of using pegs and ropes a simple portable apparatus was used in Bihar during this survey. This apparatus was a modified form of the one used by Sir John Hubback in his crop-cutting work in Bihar. It consisted of three flat iron rods which formed a right-angled isosceles triangle. The longer (hypotenuse) piece was placed on the ground with one end lying over the point located at random within the plot in the way explained above. The two shorter sides were pushed through grooves fitted at the two ends of the long piece (forming the hypotenuse) at an angle of 45° until they meet at right-angles. The procedure of pushing the two side pieces through the plants was deliberately adopted by Hubback to prevent any bias coming in through the inclusion (or exclusion) of individual plants occuring on the border line.

144. The plants enclosed in this way within the triangular iron frame were then cut and weighed. In the case of wheat, the ears were clipped off and weighed giving the weight of the crop as harvested (called for convenience the weight of 'ears'). Each cutting was put into a numbered bag and dried for a few days; the grains were then separated and weighed again giving what is called the weight of grain. In the case of gram, the pods were collected and weighed which gave the weight of pods as harvested while the grains separated from the pods after they were dried gave the weight of grain. The area enclosed within the triangular piece was 12.5 sq. feet; two cuts were obtained on either side of the hypotenuse giving two sub-cut readings for each sample-unit of area 25 sq. feet. Two such sample units (called a doublet) were cut in every selected plot.

YIELD OF WHEAT AND GRAM PER ACRE

145. The mean yield in maunds of grain per acre of wheat and gram is shown in col. (4) and col. (8) respectively of Table (6/1) separately for each sub-block the serial number of which is given in col. (1). The mean yield of wheat varied from 3·48 maunds in sub-block No. (2) to 8·40 maunds per acre in sub-block No. (1) in Shahabad district; and from 4·15 maunds in sub-block No. (2) to 10·43 maunds per acre in sub-block No. (1) in Monghyr district. In the same way, the mean yield of gram varied from 3·25 maunds in sub-block No. (6) to 6·38 maunds per acre in sub-block No. (1) in Shahabad district; and from 3·44 maunds in sub-block (No. 2) to 7.68 maunds in sub-block No (1) in Monghyr district. The variation from one sub-block to another was thus quite appreciable in each district.

THE CONVERSION FACTOR

146. The yield of both wheat and gram has naturally to be given in terms of grain after extraction from ears and pods respectively. The process of drying the crop and extracting the grains is a complicating factor and takes time. It would simplify the field work quite appreciably if the weight of the crop could be taken immediately after harvesting and later on converted into the equivalent weight of grain. It will be remembered that this is what is done in the case of paddy which is directly weighed, and this weight later converted into weight of rice (not in husk) with the help of a conversion of factor the value of which is usually accepted as two-thirds. With a view to study the possibility of using a similar method in the case of wheat and gram the cuttings in the present survey were weighed twice, as already noted, once immediately after harvesting and a second time after the extraction of the grain.

147. From the theoretical point of view the best method of finding the relation between the weight of ears of wheat or of pods of gram as harvested and corresponding equivalent weight of grain would be to fit a linear regression equation $y=b(x)$ where 'y' is the weight of grain, and 'x' the weight of crop as harvested. All available material was graduated in this way, and relevant data are given in Tables (6/1) and (6/2). In Table (6/1) col. (1) gives the serial number of the sub-block, col. (2) the number of sub-cuts; col. (3) the mean weight in maunds per acre of the crop as harvested; and col. (4), as already noted, the corresponding weight in maunds per acre of dry grain.

TABLE (6/1). RELATION BETWEEN WEIGHT OF CROP AS HARVESTED AND WEIGHT OF GRAIN OF WHEAT AND GRAM

Serial No. of sub-blocks	Wheat				Gram			
	Ears as harvested		Weight of grain in maunds per acre		Pods as harvested		Weight of grain in maunds per acre	
	No. of sub-cuts	Weight in maunds per acre	Observed values	Graduated values	No. of Sub-cuts	Weight in maunds per acre	Observed values	Graduated values
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				DISTRICT : SHAHABAD				
1	274	11 ·81	8 ·40	7 ·83	135	9 ·13	6 ·38	6 ·06
2	179	5 ·31	3 ·48	3 ·32	204	9 ·17	5 ·89	6 ·09
3	233	12 ·07	7 ·27	8 ·00	115	8 ·32	5 ·54	5 ·52
4	54	10 ·30	4 ·66	6 ·83	37	7 ·23	5 ·15	4 ·80
5	255	8 ·71	5 ·56	5 ·77	160	5 ·34	3 ·55	3 ·55
6	50	8 ·91	5 ·97	5 ·91	48	5 ·03	3 ·25	3 ·34
				DISTRICT : MONGHYR				
1	24	17 ·97	10 ·43	10 ·84	20	21 ·67	7 ·68	13 ·89
2	324	6 ·84	4 ·15	4 ·12	214	5 ·44	3 ·44	3 ·49
3	183	7 ·51	5 ·02	4 ·53	174	6 ·62	4 ·60	4 ·24
4	248	9 ·60	7 ·06	5 ·79	105	7 ·06	4 ·98	4 ·53
5	136	7 ·67	4 ·68	4 ·62	102	6 ·10	3 ·78	3 ·91
6	169	17 ·05	8 ·90	10 ·28	20	5 ·66	4 ·19	3 ·63

TABLE (6/2). COEFFICIENT OF CORRELATION AND REGRESSION BETWEEN WEIGHT OF CROP AS HARVESTED AND WEIGHT OF GRAIN OF WHEAT AND GRAM

Serial No. of sub-blocks	Wheat			Gram		
	Number of sub-cuts	Coefficient of Correlation (r)	Coefficient of Regression (b)	Number of sub-cuts	Coefficient of Correlation (r)	Coefficient of Regression (b)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			DISTRICT : SHAHABAD			
1	274	0 ·9055	0 ·7682	133	0 ·9497	0 ·7124
2	179	·8710	·6241	204	·8231	·6292
3	233	·9210	·6104	115	·9700	·6691
4	54	·9365	·5552	37	·8894	·7152
5	255	·9508	·6224	160	·9599	·6571
6	50	·9780	·6688	48	·9255	·6219
Total	1,045	·9108	·6627	697	·9234	·6643
			DISTRICT : MONGHYR			
1	24	0 ·8896	0 ·5369	20	0 ·9921	0 ·3488*
2	324	·9169	·5867	214	·9536	·5970
3	183	·9519	·6648	174	·9528	·6728
4	248	·9662	·7519	105	·9626	·7202
5	136	·9215	·6043	102	·7751	·5568
6	169	·8767	·5256	20	·8332	·7215
Total	1,084	·9108	·6031	635	·8752	·6410

* This has been omitted in forming the pooled estimate.

148. The coefficients of correlation for each sub-block are shown in Table (6/2) in which col. (1) gives the serial number of the sub-block ; col. (2) the number of sub-cuts, and col. (3) the co-efficient of correlation between the two weights for wheat in Shahabad district ; the corresponding co-efficient of regression 'b' (or factor of conversion) is given in col. (4). In the same way, the number of sub-cuts, the coefficient of correlation, and the factor of conversion are given in cols. (5), (6) and (7) respectively for grams. It will be noticed that the coefficients of correlation are all high showing a close physical connexion between the two weights.

149. The conversion factors for wheat and gram given in cols. (4) and (7) of Table (6/2), show a good deal of fluctuation from sub-block to sub-block in both districts. This may be partly due to real differences in maturity or other conditions under which the crop was harvested in different regions. Part of the observed fluctuations must also have arisen from personal factors or even careless or dishonest work on the part of the investigators. In one case, for example, for gram in sub-block (1) of Monghyr district the value of the conversion factor is extremely low, only about 0·35, while the coefficient of correlation is above +0·99 and is the highest observed value. The size of the sample is very small, only 20, but even then the results look suspicious ; and the material has been rejected in calculating pooled values of the conversion factor for each district (based on the pooled data for all sub-blocks) which are shown at the bottom of each half of the Table.

150. The pooled values are 0·663 and 0·603 for wheat and 0·664 and 0.641 for gram in the two districts Shahabad and Monghyr respectively. These pooled values were used to convert the weight of crop as harvested into corresponding weight of grain which are shown for wheat in col. (5) and for gram in col. (9) of Table (6/1). There is reasonably good agreement between these graduated values and the directly observed values of the weight of grain shown in cols. (4) and (8) for wheat and gram respectively.

151. The important point to note is that in spite of variations the values of the conversion factor are not widely different. This is encouraging as it indicates the possibility of the conversion factor being something really stable (like the factor for conversion of weight of paddy into weight of husked rice). The subject clearly deserves further study. A good plan would be to collect necessary material for this purpose at a number of selected Government farms in different parts of the province where the work can be relied upon to be done with the requisite accuracy ; the active co-operation of the Agriculture Department is essential for this purpose. Controlled experiments would also have to be made under proper supervision in cultivator's own fields for studying the factor under conditions in which the conversion actually takes place.

PRECISION OF THE RESULTS

152. I shall now examine briefly the precision of the results of crop-cutting experiments which was actually attained in this survey, and also which it is possible to attain in future.

153. *Precision of district rates of yield.* The weighted average rate of yield of wheat (ears as harvested) is 10·30±0·49 maunds per acre in Shahabad, and 9·05±0·43 maunds per acre in Monghyr. The two figures ±0·49 and ±0·43 are the standard errors of the two respective mean values. The weighted average rate of yield of gram (pods as harvested) is 7·68±0·45 maunds per acre in Shahabad and 5·93±0·54 maunds per acre in Monghyr district ; the two figures ±0·45 and ±0·54 again being the respective standard errors. It is interesting to note that standard error of the district mean yield is in every case just about half a maund per acre. Dividing the standard errors by respective mean values we also notice that the percentage standard error was 4·7% and 4·8% or just under five per cent for wheat in the two districts, and somewhat higher 5·9% and 9·1% respectively in Shahabad and Monghyr for gram. This is not unsatisfactory ; but even with the same number of cuttings the margin of error can be reduced appreciably by improving the pattern or design of the survey. No useful purpose will be served by entering into details of technicalities, but certain basic data are being given below for a general appreciation of the position and convenience of reference.

154. *Multi-stage analysis of variance.* The design of the crop-cutting experiments was of the multi-stage type. The appropriate analysis of variance for the mean yield of crop as harvested in maunds per acre is shown in Table (6/3) in which col. (1) gives the source of variation ; col. (2) the degree of freedom of the material for wheat ; col. (3) the sum of squares of deviations ; col. (4) the variance (which is obtained simply by dividing the figure in col. (3) by the corresponding figure in col. (2) ; col. (5) gives the calculated multi-stage standard deviation ; and col. (6) the corresponding coefficient of variation which is obtained

by dividing the figure in col. (5) by the general mean yield of 10·30 maunds per acre in Shahabad district and 9·05 maunds per acre in Monghyr district. Similar data for gram are given in col. (7) to (11) ; the general mean yield for conversion of standard deviations into coefficients of variation being 7·68 and 5·93 maunds per acre in Shahabad and Monghyr respectively.

TABLE (6/3). MULTI-STAGE ANALYSIS OF VARIANCE : YIELD IN MAUNDS PER ACRE OF WHEAT AND GRAM AS HARVESTED.

Source of Variation	Wheat					Gram				
				Multi-stage					Multi-stage	
	Degrees of freedom	Sum of squares	Variance	Standard deviation	Coefficient of variation	Degrees of freedom	Sum of squares	Variance	standard deviation	Coefficient of variation
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
DISTRICT : SHAHABAD										
	Mean Yield = 10·30±0·49					Mean Yield = 7·68±0·45				
Between zone-cells	29	4183·72	144·33	2·51	24·4	30	2151·06	71·70	2·22	28·9
Within cells between plots	274	4951·96	18·08	2·94	28·5	156	2065·42	13·25	2·33	30·3
Within plots between cuts	302	226·69	0·75	0·86	8·4	187	443·31	2·37	1·54	20·1
Total between cuts	605	9362·37	15·46	(3·94)	(35·3)	373	4659·79	12·49	(3·53)	(46·0)
DISTRICT : MONGHYR										
	Mean Yield = 9·05±0·43					Mean Yield = 5·93±0·54				
Between Zone-cells	27	2649·78	98·13	2·09	23·1	26	2319·81	89·24	2·3	42·3
Within cells between plots	239	3687·52	13·41	2·70	29·8	128	2125·87	16·62	2·59	43·7
Within plots between cuts	267	259·44	0·96	0·98	10·8	155	513·84	3·32	1·83	30·9
Total between cuts	533	6595·74	12·37	(3·53)	(38·9)	309	4959·52	16·04	(4·00)	(67·5)

155. *Design of crop-cutting experiments:* Without entering into technicalities it may be mentioned here that the values of the standard deviations given in col.(5) and (10) indicate that nothing is to be gained by increasing the number of sample-cuts within each plot nor by increasing the number of plots within each zone-cell. The precision of the results would be increased by scattering the sample-cuts as widely as possible over different zone-cells rather than on a large number of plots within particular zone-cells, and over plots rather than on cuts within plots. This is just what one would expect as the more widely the sample-cuts are scattered the more representative are likely to be the result.

156. The multi-stage standard deviations 'between zone-cells' and 'within zone-cells between plots' are, roughly, of the same order for both crops showing that the fluctuations in yield from one zone-cell to another or from one plot to another within zone-cells are about the same for both wheat and gram in each district. The standard deviation within plots is, however, greater for gram showing that the yield of this crop is more variable from one part to another of the same plot. But the important point to note is that the standard deviations for both crops are, roughly, of the same order so that the same design of survey would serve quite well for both crops.

POSSIBILITIES OF IMPROVEMENT IN PRECISION

157. *District rates of yield:* It will be noticed from col.(6) of table (6/3) that the overall or total standard deviation is 3·94 maunds per acre for wheat for the Shahabad material and 3·53 maunds per acre for the Monghyr material. The corresponding values for gram are 3·53 and 4.00 maunds per acre in the two districts respectively. We may, therefore, adopt say 4 maunds per acre as a convenient value of the standard deviation for purposes of illustration.

If 256 cuts are collected purely at random (in the form of uni-stage sampling) from all over the district then the standard error of the mean yield of the district would be reduced to one quarter of a maund per acre so that in 19 cases out of 20 the observed mean value would not be out by more than half a maund per acre. We can express the same results in terms of percentage margin of error. From col.(6) of Table (6/3) we notice that the coefficient of variation for wheat is practically same, 38·25% and 38·89% in Shahabad and Monghyr respectively. For gram the values are 45·96% in Shahobad and 67·45% in Monghyr. The proportionate variability is thus somewhat higher for gram as compared to wheat. For purposes of illustration we may adopt say 40% for wheat and 64% for gram. With 256 cuttings spread purely at random all over the district the mean yield would have a standard error of 2·5% for wheat and 4% for gram.

158. *Margin of error of provincial rates of yield.* The use of multi-stage (instead of simple random or uni-stage) sampling is, however, usually more convenient in practice although the margin of error is somewhat larger. With the values of the standard deviation (or coefficient of variation) for different stages given in Table (6/3) it is possible to calculate the margin of error for different patterns of sampling. The coefficient of variation for wheat, it will be noticed from col. (6) of table (6/3), lies between 23·1% and 29·8%. If the fluctuation of yield in other districts are of the same order then we may use 25% as a convenient value for purposes of illustration. For gram the values lie between 28·9% and 43·7%, and we may adopt 40% for numerical calculations. If 4 cuts were obtained from each of 625 zone-cells then with 2500 cuts the standard error of the mean yield for the province as a whole would be about 1·1% for wheat and 1·8% for gram which would be adequate for all practical purposes. We thus find that 2500 or 3000 sample-cuts distributed in the proper way would enable the provincial mean yield of wheat being determined with a standard error of something like one per cent, and of gram with a margin of error of 2% or less. As the physical uncertainty itself is of the same magnitude nothing would be gained by trying to reduce the sampling error below this limit.

159. The actual pattern of sampling has, however, to be setteld from consideration of cost and practical convenience of field work. The values of the standard deviations and coefficients of variation at different stages of sampling given in Table (6/3) in conjunction with considerations of cost and convenience enable the sampling technique being settled on a scientific basis. The values of the standard deviations obtained during the present survey show, however, that working with handy cuts of the kind used here there would be no difficulty in determining the district or provincial mean rates of yield with all necessary or desired precision.

EFFECT OF PREMATURE HARVESTING

160. A preliminary study was made of the effect on the rate of yield due to harvesting the crop before it was fully mature. One plot under wheat at Sasaram and another plot under gram in Mohania were selected for this purpose. Within these plots a grid of size 20′×20′ was marked out; and within this enclosure 10 plants were picked up at random at intervals of 3 days until the plot was harvested by the owner. Weights were taken of individual ears from which the mean weight of ears at harvesting was calculated. Later on the crop was dried and the weight of grain was also determined, and the ratio of the weight of grain to weight of ears of wheat (or pods of gram) was calculated.

161. Table (6/4) shows the results of the work on wheat in Sasaram; col.(1) gives the date of collecting the plants; col.(2) the number of plants; col.(3) the number of ears. The mean weight of the crop in tolas per ear as harvested is given in col.(4); the mean weight of grain in tolas per ear in col.(5); and the mean ratio of the weight of grain of wheat to weight of ears

TABLE (6/4). WEIGHTS (IN TOLAS PER EAR) OF WHEAT.

Date of cutting plants from field	Number of plants of wheat	Number of ears of wheat	Mean weight in tolas per ear with standard error		Ratio of weight of grain to weight of crop as harvested
			crop as harvested	grain after extraction	
(1)	(2)	(3)	(4)	(5)	(6)
28 March 1944	10	71	0·132±0·008	0·057±0·001	0·372±0·054
31 March 1944	10	75	0·163±0·012	0·064±0·002	0·432±0·083
3 April 1944	10	73	0·155±0·009	0·090±0·003	0·579±0·132
6 April 1944	10	71	0·134±0·007	0·067±0·002	0·503±0·091
9 April 1944	10	66	0·122±0·006	0·070±0·003	0·571±0·176
12 April 1944	10	55	0·127±0·006	0·077±0·003	0·591±0·146

as harvested is shown in col.(6). In the same way results for work on gram are shown in Table (6/5). As before, cols.(1) and (2) give the date of cutting and number of plants respectively. The weight of pods as harvested in tolas per plant is given in col.(3); the weight of gram (after extraction from pods) in tolas per plant in col.(4); and the ratio of the above two weights in col.(5).

TABLE (6/5). WEIGHTS (IN TOLAS PER PLANT) OF GRAM.

Date of cutting of plants from field	No. of plants	Weight in tolas per plant of		Ratio of weight of pods to weight of gram
		pods as harvested	grain after extraction	
(1)	(2)	(3)	(4)	(5)
29 March 1944	10	10·0	6·5	0·65
3 April 1944	10	10·0	6·5	0·65
6 April 1944	10	6·5	4·0	0·62
9 April 1944	10	6·5	4·0	0·62

162. It will be noted that the weight of wheat ears decreased with increasing maturity while the grain gained in weight. The ratio of weights therefore increased with increasing maturity. In the case of gram both weights decreased with increasing maturity so that the ratio was more or lesss steady.

163. The above experiment was undertaken to see whether it was possible to make any corrections for premature harvesting of the crop. As pointed out in another section, one great difficulty of working with a comparatively smaller number of investigators in a crop-cutting survey is the fact that the investigators have to move rapidly from one place to another in order to cover the whole province in an adequate manner. It is therefore often inevitable that an investigator should reach a village when the crop on the plots (which were selected beforehand at random for crop cutting work) was not yet ripe for harvesting. If the investigator has to wait until the crop becomes ripe, a good deal of time may be wasted. On the other hand, if he leaves the village without collecting the crop, a large number of gaps would occur in the sample. In this situation, if it were possible to cut the crop even when it was not fully mature and use suitable adjustments to obtain the estimated crop at maturity, then the efficiency of crop-cutting work could be increased appreciably. The results obtained in the present preliminary study are not adequate for this purpose, but show that it may be worth while exploring the possibilities in this direction.

MEAN RATES OF YIELD PER ACRE OF WHEAT AND GRAM

164. We may now consider the full material comprising all cuts for which the weight of the crop as harvested was available. The data are given in Table (6/6) in which col.(1) shows the serial number of the sub-block; col.(2) the number of sub-cuts of wheat ; and col.(3) the mean weight in maunds per acre. These were then converted into weight (in maunds per acre) of dry grain by multiplying by appropriate conversion factors based on linear regression; these values are shown in col.(4). In the same way, the number of sub-cuts of gram as harvested is shown in col.(5); the corresponding weight in maunds per acre in col.(6); and the converted weights (in maunds per acre) of dry grain in col.(7).

165. Results of crop-cutting work in the special camps at Mohania and Sahpur in district Shahabad and at Gogri and Kanui in district Monghyr have not been given in Table (6/6) ; owing to their highly local character these values have neither been taken into consideration in calculating the district averages.

166. It will be noticed from col.(3) of Table (6/6) that there were considerable variations in the yield from sub-block to sub-block, that is, from one geographical region to another within each district. In Shahabad the yield of wheat (as harvested) varied from 7·60 maunds per acre in sub-block (2) to 12·42 maunds per acre in sub-block (1). In Monghyr the variation was even larger; from 6.84 maunds per acre in sub-block (2) to 17·97 maunds per acre in sub-block (1).

167. The variation was also quite large for gram (as pods at the time of harvesting) from 4.62 maunds per acre in sub-block (6) to 8·74 maunds per acre in sub-block (1) in Shahabad district as shown in col.(6) of Table (6/6). Omitting a very high value of 21·67 maunds per acre in sub-block (1) in Monghyr district (which is unreliable as it was based on only 24 sample-cuts), the mean yield of gram was more steady and remained between 5.26 and 6·80 maunds per acre in the different sub-blocks.

Table (6/6). Mean yield in maunds per acre of wheat and gram

Serial No. of sub-blocks	Wheat			Gram		
	No. of sub-cuts	Weight in maunds per acre: ears as harvested	Weight in maunds per acre: converted into grain	No. of sub-cuts	Weight in maunds per acre: pods as harvested	Weight in maunds per acre: converted into grain
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			District : Shahabad			
1	296	12·42	8·23	148	8·74	5·80
2	270	7·60	5·04	248	8·56	5·68
3	236	11·86	7·86	116	7·97	5·29
4	64	8·95	5·93	37	7·23	4·80
5	348	8·36	5·54	226	5·22	3·47
6	52	8·82	5·85	56	4·62	3·07
			District : Monghyr			
1	24	17·97	10·81	20	21·67	13·89
2	324	6·84	4·12	115	5·26	3·37
3	183	7·51	4·53	179	6·25	4·01
4	268	9·62	5·80	117	6·80	4·36
5	148	7·01	4·23	115	5·73	3·67
6	169	17·05	10·28	20	5·66	3·63

DISTRICT OUTTURN OF CROP

168. Having obtained estimates of crop acreage and also of yield per acre it was easy (in theory) to calculate the total outturn. Multiplying the estimated acreage by the estimated rate of yield per acre for each sub-block, the outturn for each sub-block was directly obtained. From these the district totals were built up by straightforward addition, and are shown in Table (6/7). The serial number of the sub-block is given in col. (1); the total geographical area of the sub-block in thousand acres in col.(2); the sample estimate of the area of wheat (in thousand acres) in col. (3); and the yield of grain in maunds per acre of wheat in col.(4). Multiplying these two figures we get the total outturn of wheat in thousand maunds of grain in col. (5). In the same way, the area under gram is given in col. (6); the yield of grain in maunds per acre in col.(7); and the outturn of gram in thousand maunds of grain in col.(8) of table (6/7) for each sub-block separately.

169. Adding the outturn for individual sub-blocks we get in col. (5) of table (6/7) the total outturn of wheat grain of 1947·1 thousand maunds in Shahabad and 1832·9 thousand maunds in Monghyr district. Dividing these figures by the corresponding sample estimates of crop acreage from col. (3), namely, 300·35 and 284·08 thousand acres respectively, we get the adjusted average rates of yield of wheat grain of 6·48 and 6·45 maunds per acre for Shahabad and Monghyr respectively as shown within brackets in col. (4) of Table (6/7). In the same way, we get the total outturn of gram (extracted from pods) of 1680·6 and 822·4 thousand maunds for Shahabad and Monghyr as shown in col. (8) of Table (6/7). The corresponding adjusted average rate of yield for gram was 4·98 maunds per acre in Shahabad and 4.25 maunds per acre in Monghyr.

OFFICIAL ESTIMATES OF YIELD AND OUTTURN

170. *District figures.* It is difficult to make any comments on the above figures. Rates of yield and outturn are not given separately for districts in official publications of the Agriculture Department but can be calculated in the way explained below. In the fourth wheat forecast the condition factor for outturn is given as 69 per cent in Shahabad and 77 per cent in Monghyr district. It is also stated that the standard rate of yield is 12 maunds per acre; multiplying this by the respective condition factors of 69 per cent and 77 per cent one gets the calculated rates of yield of 8·28 maunds per acre in Shahabad and 9·24 maunds per acre in Monghyr shown within brackets (to indicate that these are derived figures) in col. (4) of Table (6/7). Multiplying these calculated rates by corresponding official acreage figures of 263·8 and 219·8 thousand acres shown in col. (3) of Table (6/7), we get the calculated outturn in thousands of maunds of 2184 and 2051 in Shahabad and Monghyr respectively shown within brackets in col. (5) of Table (6/7).

TABLE (6/7). CALCULATED OUTTURN IN THOUSANDS OF MAUNDS OF GRAIN OF WHEAT AND GRAM

Serial No. of sub-blocks	Geographical area in thousand acres	Wheat			Gram		
		area in thousand acres	yield in maunds per acre	Outturn in thousand maunds	Area in thousand acres	yield in maund per acre	Outturn in thousand maunds
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			DISTRICT : SHAHABAD				
1	463 ·6	65 ·84	8 ·23	541 ·9	83 ·90	5 ·8	486 ·6
2	445 ·5	62 ·32	5 ·04	314 ·1	67 ·54	5 ·6	383 ·6
3	429 ·0	50 ·71	7 ·86	398 ·6	61 ·52	5 ·2	325 ·4
4	518 ·1	34 ·87	5 ·93	206 ·8	48 ·34	4 ·8	232 ·0
5	431 ·7	67 ·67	5 ·54	375 ·0	49 ·08	3 ·4	170 ·3
6	508 ·5	18 ·92	5 ·85	110 ·7	26 ·95	3 ·0	82 ·7
	2796 ·4	300 ·35	(6 ·48)	1947 ·1	337 ·33	(4 ·98)	1680 ·6
Calculated (official sources)		263 ·8	(8·28)	(218 ·4)	(282 · 3)	(9 ·71)	(27 ·41)
			DISTRICT : MONGHYR				
1	393 ·6	3 ·95	10 ·84	42 ·8	5 ·64	13 ·89	78 ·3
2	359 ·2	15 ·76	4 ·12	64 ·9	17 ·38	3 ·37	58 ·6
3	424 ·1	52 ·33	4 ·53	237 ·0	72 ·56	4 ·01	291 ·0
4	403 ·0	72 ·70	5 ·80	421 ·7	52 ·03	4 ·36	226 ·9
5	390 ·2	60 ·48	4 ·23	255 ·8	33 ·03	3 ·67	121 ·2
6	399 ·7	78 ·86	10 ·28	810 ·7	12 ·79	3 ·63	46 ·4
	2369 ·8	284 ·08	(6 ·45)	1832 ·9	193 ·43	(4 ·25)	822 ·4
Calculated (official sources)		219 ·8	(9 ·24)	(2031)	210 ·5	(9 ·97)	(2099)

171. For gram the calculation is a little more involved. The outturn for the province as a whole is given as 541,800 tons, and the total area under gram as 1,459,600 acres in the forecast of *rabi* crops issued on the 26th April 1944 by the Director of Agriculture. Dividing 541,800 tons by 1,459,600 acres one gets 10·1 maunds per acre as the provincial rate of yield. Dividing by the condition factor of 78 per cent for the province as a whole (as given in the *rabi* forecast) the standard yield is apparently 12·95 maunds per acre. The condition factors for gram are given (in the *rabi* forecast) as 75 per cent and 77 per cent for Shahabad and Monghyr respectively; multiplying 12·95 maunds per acre by these two factors the calculated rates of yield are 9·71 and 9·97 maunds of grain per acre for Shahabad and Monghyr respectively and are shown within brackets in col. (7) of Table (6/7). Finally, multiplying the above rates of yield by corresponding official estimates of area under gram in thousand acres of 282·3 and 210·5 we get the calculated outturn of 2741 and 2099 thousand maunds of gram for Shahabad and Monghyr respectively which are shown within brackets in col. (8) of Table (6/7).

172. *Provincial figures :* Official figures relating to rates of yield and outturn are, however, of extremely doubtful value. For example, for wheat the estimated acreage for the province as a whole in 1943–44 is given as 1,220,800 acres and the outturn as 452,700 tons in the fourth wheat forecast issued by the Director of Agriculture; dividing the latter by the former we get the rate of yield as 10·09 maunds per acre (which it may be noted incidentally is practically the same as the yield of gram). The condition factor for the whole province is given as 76 per cent. (We have checked this by direct calculation; it is the average value based on the condition factor for each district weighted by the respective crop acreage). Dividing 10·09 maunds per acre by this provincial condition factor of 76 per cent we obtain a calculated standard rate of yield of 13·28 maunds per acre instead of the quoted standard figure of 12 maunds per acre. No explanation has been given of this discrepancy. In fact the position has been made even more confusing by a remark added in a footnote in the fourth wheat forecast that the Departmental estimate of the provincial condition factor is 85 per cent. Dividing the calculated provincial yield rate of 10·09 maunds per acre by this 'departmental' condition factor 85 per cent one gets a standard yield of 11·87 maunds per acre which is very nearly the same as the quoted figure of 12 maunds per acre. But this does not explain anything as the provincial outturn was probably obtained by using the condition factor of 85%. The real point is how was the 'departmental' estimate of 85 per cent obtained? Nothing has been stated in this regard.

173. To study this question a little further the provincial rates of yield and the standard yield for wheat, barley and gram for the four crop seasons from 1940–41 to 1943-44 were calculated from information given in official publications. The results of such calculations are shown in Table (6/A) in which col. (1) gives the crop season, col. (2) the area under each crop in acres in the province as a whole and col. (3) the total outturn in tons both as stated in official reports. Dividing the outturn figure in col.(3) by the corresponding acreage figure in col. (2) we get the calculated provincial rate of yield in maunds per acre shown in col.(4). The provincial condition factors as given in official reports are shown in col. (5). Dividing the calculated yield rate in col. (4) by the corresponding condition factor in col. (5) we get the calculated 'standard' rates of yield per acre shown in col. (7). In certain instances what are called 'departmental' estimates are also given of the condition factor; these are shown in col. (6). Using these condition factors it is possible to obtain a second set of calculated standard rates of yield which are shown in col. (8) headed 'departmental' rates.

Table (6/A). Calculated rates of yield based on official sources

Crop season	Area under crop in acre (official)	Produce in tons (official)	Yield in maunds per acre (calculated)	Official condition factor		Standard yield in maunds per acre	
				Provincial average	Departmental average	Provincial	Departmental
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Wheat				
1940-41	10,96,400	4,05,500	10·07	84	—	11·99	—
-42	13,00,000	4,84,600	10·14	85	—	11·93	—
-43	12,80,000	5,81,200	12·36	91	—	13·58	—
-44	12,20,800	4,52,700	10·09	76	85	13·28	11·87
			Barley				
1940-41	12,98,200	4,46,000	9·35	87	—	10·75	—
-42	12,84,300	4,46,300	9·46	88	—	10·75	—
-43	12,69,100	4,16,000	8·93	70	83	12·76	10·76
-44	12,79,100	4,95,000	10·54	82	99	12·85	10·65
			Gram				
1940-41	14,44,300	4,67,700	8·81	82	—	10·74	—
-42	14,48,600	5,29,500	9·95	85	—	11·71	—
-43	14,46,500	4,68,000	8·81	69	82	12·77	10·74
-44	14,59,600	5,41,800	10·10	78	94	12·95	10·74

174. For 1940–41 and 1941–42 figures of acreage (col.2),and produce(col.3) and condition factor (cols. 5 and 6)are taken from Appendix III and V of the Season and Crop Report, Bihar, 1941–42. For 1942–43 and 1943–44, figures are taken from the 'Fourth forecast of the wheat crop of Bihar 1943–44' and 'Forecast of the *Rabi* crop of Bihar 1943–44' dated 26th April 1944 isssued by the Department of Agriculture. The standard or normal rate of yield of wheat is given as 12 maunds per acre in the 'Fourth forecast of the wheat crop in Bihar 1943–44.'

175. It will be noticed that for wheat the "standard" yield was 11·99 and 11·95 maunds per acre in 1940-41 and 1941-42 but jumped to 13·58 maunds in 1942-43 and 13·28 maunds per acre in 1943-44. For barley the standard yield was 10·75 maunds per acre in 1940-41 and 1941-42 but suddenly increased to 12·76 and 12.85 maunds per acre in 1942-43 and 1943-44 respectively. For gram the standard rate was 10·74 maunds per acre in 1940-41, 11·71 maunds in 1941-42, 12·77 maunds in 1942-43 and 12·95 maunds in 1943-44. Such sudden changes in the "standard" yield is certainly curious. It would be also noticed that the so-called "departmental" estimates of the condition factor, given in col. (6) of Table (6/A), bring down the "standard" yield of wheat in 1943-44 to 11·87 maunds per acre as already noted, and of barley and gram to about 10¾ maunds per acre in 1942-43 and 1943-44.

176. In view of the conflicting nature of estimates based on official publications I made personal enquiries from Mr. A. P. Cliff of the Bihar Agriculture Department. In his D. O. No. 228 dated Patna the 16th July 1944 he wrote :

"There have been no crop cutting experiment by this department on wheat and gram. The normal outturn of wheat is taken to be 12 maunds per acre over the whole province vide p. 16 para 58 of the Crop Statistics Manual of Bengal 1908. The provincial normal outturn of gram is similarly taken as $10\frac{3}{4}$ maunds per acre. I am unable to give you any 'evidence' for these figures and doubt very much if any, worth the name, exists."

It is clear that standard rates of yield used by the Agriculture Department are purely conventional and have no basis in experimental observations. We have already noted that official acreage figures relating to "mixed crops" are extremely ambiguous. Comparison of our figures with official estimates is, therefore, not possible.

SPECIAL STUDIES

177. *Mixed crops.* Certain special studies were also attempted in connexion with the determination of yield and total outturn. For example, a preliminary study was made of the yield per acre of wheat from plots which were sown with different proportions of wheat in mixture with other crops ; the results have been discussed in the next section 7.

178. One point may however be noted here. In the present survey the crop acreage given is the net or effective area which would have been covered by a crop if it had been sown "pure" by itself, that is, without any mixture with other crops. The rates of yield per acre is, however, based on the gross or total geographical area covered by all crops in the mixture. To calculate the outturn by multiplying the above two figures would obviously lead to under-estimation. Official crop acreages are probably net areas (although nothing is known definitely about this), while standard rates of yield are presumably based on gross area; it is likely, therefore, that official outturn figures are also under-estimates.

179. *Variance function and border effect.* A certain amount of material was collected to study in what way the variance of yield decreased with the increase in the size of the sample-cut ; but the material was too meagre to give any clear indication. There is urgent need of a critical study of the variance function and also of the border effect, that is, whether any bias is being introduced or not due to inclusion (or exclusion) within the sample-cuts of plants occurring on the border of the cuts. This bias, if it exists, is naturally likely to be comparatively more important in cuts of a small size as the proportion of the perimeter to the total area of the cut is large.

180. *Eye-estimation of yield per acre.* In Bengal in 1943 and in previous years very encouraging results had been obtained in controlled experiments in eye-estimation by physical examination of the yield per acre of paddy and jute crops. A number of investigators were found capable of doing this with close approximation to results based on direct crop-cutting work. An attempt was made to carry out a similar experiment on wheat in Bihar, but owing to shortness of time the investigators could not be given necessary training and the material collected was not of much value.

181. It will be noticed there are many problems requiring critical study in connexion with the determination of yield per acre and outturn of *rabi* crops in Bihar. A good begining has, however, been made in the pilot survey in 1943-44 which must be followed up in future.

SECTION 7. THE PROBLEM OF 'MIXED' CROPS

182. In the *rabi* season in Bihar the outstanding difficulty was undoubtedly the occurence of mixed crops of many different types. Wheat, gram and barley are usually sown in mixtures of two, three or more crops ; and plots under a single crop are rare except for *arhar*. It is therefore necessary to enquire what is the exact meaning of the area sown under mixed crops, and how is one to estimate the area separately for each component crop.

TWO WAYS OF OCCURRENCE OF 'MIXED' CROPS

183. So far as an individual plot is concerned the word 'mixed' may have two different meanings. For example, one portion say, 4-anna proportion of a plot may be entirely under wheat ; another portion say, 6-anna, entirely under barley ; and the remaining portion, say 6-anna entirely under gram. In such cases there is, of course, no ambiguity as to how much area has to be allotted to each individual crop.

184. There is however a second way in which mixed crops can occur. For example, the sample plot may be sown with a mixture of wheat, barley and gram scattered all over the plot. In this case, in one sense, each crop covers the whole or 16-anna of the plot. Should then each crop be credited with the full 16-anna of the land ? The total anna estimate would then obviously total up to $16+16+16=48$ anna or three times the actual geographical area of the plot itself. A second method would be to make an estimate of the net anna-proportion of the land which would have been covered by, say, wheat if it had been sown separately from the other two crops. In this method the net or effective proportion of land

under each crop has to be estimated, and the total for each plot (inclusive, of course, of land not under cultivation) would be invariably 16-anna or exactly equal to the geographical area.

VARIATIONS IN COMPOSITION OF 'MIXED' CROPS

185. A complication is, however, introduced by the fact that when the same two or the same three crops are sown mixed in the second way (that is, covering the same identical portion of the plot), the intensity or relative proportion of the different crops do not remain the same but vary widely from region to region, and even from plot to plot. Results of certain sample studies for wheat are shown in Tables (7/1) and (7/2) in which the anna composition of different types of mixed crops are shown based on the analysis of records relating to 16,564 plots in Shahabad district and 16,724 plots in Monghyr district. In Table (7/1) the first three columns show the anna composition of the crop under three heads, wheat, gram, and other crops respectively. Thus 16-0-0 indicates plots entirely under wheat, 0-16-0 means plots entirely under gram, or 0-0-16 plots entirely under other crops (which themselves may be mixed). In the same way, 1-5 anna (under wheat), 11-15-anna (under gram) means a mixed crop of wheat and gram in which the anna proportion of wheat lies between 1-anna and 5-anna and the anna-proportion of gram between 11-anna and 15-anna and so on. Col. (4) shows the number of plots of each type occurring in the sample in Shahabad district; col. (5) gives the percentage of all plots; and col. (6) the percentage in terms of cultivated plots only. Similar information is given in cols. (7), (8) and (9) for Monghyr district.

186. It is clear from Table (7/1) that plots under wheat or gram alone are rather rare being only 3·2 per cent of all plots or 6·2 per cent of cultivated plots in Shahabad, and 3·1 per cent of all plots and 8·0 per cent of cultivated plots in Monghyr. Wheat and gram in mixture occur in 12·9 per cent of cultivated plots in Shahabad, and in 15·0 per cent of such plots in Monghyr; wheat occurs in mixture with gram and other crops in 22·1 per cent of cultivated plots in Shahabad and in 15·7 per cent of such plots in Monghyr; finally, wheat occurs in mixture with crops other than gram in 5·4 per cent of cultivated plots in Shahabad and 8·1 per cent of such plots in Monghyr. Gram occurs with crops other than wheat in Shahabad in 14·5 per cent and in Monghyr in 7·9 per cent of cultivated plots. Wheat thus occurs in mixture with some crop or other in Shahabad in 40·4 per cent, and by itself only in 3·8 per cent of all cultivated plots; in Monghyr wheat occurs in mixture in 38·8 per cent, and by itself in only 5·2 per cent of all cultivated plots. In the same way, gram occurs in mixture with other crops in 49·5 per cent and by itself in 2·4 per cent of cultivated plots in Shahabad; and in mixture in 38·6 per cent and by itself in 2·8 per cent of cultivated plots in Monghyr. It will be noticed that the composition of mixed crops of wheat and gram thus differs quite appreciably from plot to plot.

187. The position so far as a particular component crop, say wheat, is concerned can be seen better if the material is tabulated in a different way as in Table (7/4) in which col. (1) shows the type of composition in terms of anna proportion of intensity of cultivation of wheat alone, namely, plots having effective anna-proportion of wheat between 1-anna and 5-anna, 5.3-anna (*i.e.* roughly equal mixture of 3 crops), between 6-anna and 10-anna, between 11-anna and 15-anna, and 16-anna or entirely under wheat; col. (2) gives the number of plots observed in each category; and col. (3) the same information expressed in the term of percentages of the total number of plots studied. It will be seen from col. (3) of Table (7/4) that in Shahabad district 6·5 per cent of plots had usually between 1-anna and 5-anna effective intensity of cultivation in mixed crops; 50·0 per cent had one third (or 5·3-anna) effective composition of wheat in mixed crops of wheat, gram and some other crop; about 20·4 per cent of plots had from 6-anna to 10-anna or a medium effective intensity of cultivation of wheat; and 14·3 per cent of plots had from 11-anna to 15-anna or a heavy proportion of wheat in mixed crops. Finally, the proportion of plots entirely under wheat (among all plots having wheat) was only 8·8 per cent. In Monghyr district the position was broadly similar; 8·1 per cent of plots had an effective anna-composition of 5-anna or less under wheat; 35·7 per cent had one third (or 5·3-anna) composition; 34·3 per cent between 6-anna and 10-anna composition; 10·1 per cent with between 11-anna and 15-anna composition of wheat; while only 11.8 per cent of plots were entirely under wheat.

GROSS AND NET AREA UNDER INDIVIDUAL CROPS

188. This is the actual position. How then is the total area under wheat (or other crops) to be estimated? Our first reaction naturally was to adopt existing standards in use in the Bihar Agricultural Department so that our sample estimates would be directly

TABLE (7/1). DISTRIBUTION OF PLOTS UNDER WHEAT AND GRAM BY TYPE OF MIXTURE OF CROPS.

Anna-composition of type of mixed crop			Shahabad			Monghyr		
Wheat	Gram	Others	Number of plots	Percentage of plots to total	Percentage of plots to cultivated	Number of plots	Percentage of plots to total	Percentage of plots to cultivated
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			(1) Single Crop (wheat or gram)					
	0	0	322	2·0	3·8	341	2·0	5·2
0	16	0	200	1·2	2·4	185	1·1	2·8
Total			522	3·2	6·2	526	3·1	8·0
		e	(2) Wheat and Gram					
1—5	11—15	—	% 94	0·6	1·1	102	0·6	1·6
6—10	6—10	—	607	3·6	7·3	763	4·6	11·6
11—15	1—5	—	377	2·3	4·5	115	0·7	1·8
Total			1078	6·5	12·9	980	5·9	15·0
			(3) Wheat and Others (not gram)					
1—5	—	11—15	148	0·9	1·8	133	0·8	2·0
6—10	—	6—10	149	0·9	1·8	226	1·4	3·4
11—15	—	1—5	153	0·9	1·8	175	1·0	2·7
Total			450	2·7	5·4	534	3·2	8·1
			(4) Gram and Others (not wheat)					
—	1—5	11—15	370	2·2	4·4	156	0·9	2·4
—	6—10	6—10	481	2·9	5·8	264	1·6	4·0
—	11—15	1—5	361	2·2	4·3	97	0·6	1·5
Total			Œ 1212	7·3	14·5	517	3ö1	7·9
			(5) Wheat, Gram and others					
5·3	5·3	5·3	1847	11·2	22·1	1029	6·2	15·7
			(6) Other crops (not wheat or gram)					
—	—	16	3255	19·6	38·9	2975	17·8	45·3
Cultivated plots			8364	50·6	100·0	6561	39·3	100·0
			(7) Not cultivated					
0	0	0	8200	49·5	—	10163	60·7	—
All plots			16564	100·0	—	16724	100·0	—

TABLE (7/2). SUMMARY DISTRIBUTION OF PLOTS UNDER WHEAT AND GRAM BY TYPE OF MIXTURE OF CROPS

	Shahabad			Monghyr		
	Number of Plots	Percentage of plots to total	Percentage of plots to vultivated	Number of Plots	Percentage of plots to total	Percentage of plots to cultivated
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 Single crop (wheat or gram)	522	3·2	6·2	526	3·1	8·0
2 Wheat and gram	1078	6·5	12·9	980	5·9	15·0
3 Wheat and others	450	2·7	5·4	534	3·2	8·1
4 Gram and others	1212	7·3	14·5	517	3·1	7·9
5 Wheat gram and others	1847	11·2	22·1	1029	6·2	15·7
6 Other crops	3255	19·6	38·9	2975	17·8	45·3
Total cultivated	8364	50·5	100·0	6561	39·3	100·0
(Not cultivated)	8200	49·5	—	10163	60·7	—
All plots	16564	100·0	—	16724	100·0	—

comparable with departmental figures. I made enquiries in the matter, and also requested Rai Bahadur S. P. Sinha to discuss the subject with the officers of the Agricultural Department and to ascertain what methods were actually being used for estimating the area allotted to individual crops in the case of plots with mixed crops. No definite information could be secured but I was given to understand that the Agricultural Department was not using any standard method for this purpose. The area figures for individual *mauzas* are supposed to be supplied by local officers, and nothing was known as to how the acreage for individual crops was actually estimated.

189. The Settlement Reports also failed to make the position clear. In the Gaya Report (1911-1918) by Mr. E. L. Tanner, I.C.S., it is stated in para 136, p. 81 :—

> "If in any particular plot wheat and barley mixed were recorded as having been grown, the area was proportionately divided between them".

This, of course, is the method of using net or effective acreages. In the same Report in para 140, p. 82, it is, however, again stated :—

> "In the first year's area an attempt was made to reduce all the results to a 16-anna crop by estimating, what the actual crop was in terms of a bumper crop (16-anna). The result was so inaccurate that this was dropped and only actual results recorded...... and in giving the results of the experiment only actual outturns have been given".

The attempt to make adjustments for the variation in the yield due to differences in the effective composition of wheat in mixed crops was thus apparently abandoned.

GROSS AREA WITH ANNA COMPOSITION

190. In the absence of any standard method the Field Branch was provisionally instructed to try to estimate the net or *effective* anna proportion of the plot covered by individual crops. On the 25th and 26th March 1944, when I was at the special camp at Bikramgunj in Shahabad district, this question was discussed in considerable detail at a conference with the technicians and the supervising staff of the Field Branch. It was agreed that the proper course would be to record both (*a*) the gross area as well as (*b*) the estimated effective composition of each component crop either in anna terms or in certain standard broad categories such as 'low wheat and heavy gram', 'medium wheat and medium gram', etc. By direct experiment we found that there was no difficulty in classifying plots under mixed crops in such broad categories.

YIELD RATES FOR MIXED CROPS

191. It was also decided that if crop-cutting experiments could be carried out on a sufficient number of plots of each of the different types of mixed crops then it would be possible to determine how the rate of yield per acre of each individual crop depended on its composition in the mixture. These rates of yield could be then used to calculate the total production or outturn from plots in different categories. It was too late in the season to carry out this programme in full but a certain amount of material was secured from which results of considerable interest have been extracted.

192. From available crop-cutting records an attempt was made to classify the plots under mixed crops by the proportion or anna-estimate of effective composition of wheat. Unfortunately, the anna composition of the crop was, as a rule, not mentioned in the crop-cutting record itself; the serial number of the grid (from which the plot was selected) was also not given. The serial number of the plot was, however, available in the original *khasra* or field records for area survey; it was possible, therefore by searching through these lists, to trace the corresponding effective anna-composition of wheat in the given plot as recorded at the time of the area survey in the field *khasra* books. The procedure was extremely laborious. Relevant information was however traced for 189 plots in Shahabad district and for 247 plots in Monghyr district.

193. The yield of wheat in maunds per acre was calculated separately for plots in different anna-categories. The results are shown in Table (7/3) in which the anna-composition of the plots is given in col. (1) in two-anna intervals, that is, with estimated compositions of 1 and 2-anna, 3 and 4-anna, 5 and 6-anna, ...15 and 16-anna under wheat. The number of plots in each category is given in col. (2); and the observed mean yield in maunds of wheat grain per acre is shown in col. (3). In the same way, the number of plots in each class in Monghyr district is shown in col. (5); and the observed yield of wheat in col. (6).

TABLE (7/3). RELATION BETWEEN YIELD IN MAUNDS OF GRAIN PER ACRE AND ANNA PROPORTION OF INTENSITY OF CULTIVATION OF WHEAT IN MIXED CROP

Anna proportion of intensity of cultivation of wheat in mixed crops (sample plots)	District Shahabad			District Monghyr		
	No. of plots in sample	Mean yield in maunds per acre		No. of plots in sample	Mean yield in maunds per acre	
		observed	graduated		observed	graduated
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1-anna to 2-anna	9	4·0	4·3	11	5·6	4·2
3-anna to 4-anna	17	5·7	4·9	23	4·5	4·8
5-anna to 6-anna	30	5·1	5·5	24	4·2	5·3
7-anna to 8-anna	32	5·7	6·1	55	5·4	6·0
9-anna to 10-anna	35	7·1	6·6	35	6·6	6·6
11-anna to 12-anna	19	6·9	6·9	35	6·4	7·2
13-anna to 14-anna	10	7·1	7·2	14	9·6	7·8
15-anna to 16-anna	37	7·5	7·4	50	8·1	8·4
Total	189	6·3	6·3	247	5·5	5·5
All sub-cuts	1266	6·21	—	1112	5·58	—

194. It will be noticed that, on the whole, the yield increased, as one would expect, with increasing intensity of anna-composition. A second degree parabola was used for graduating the material for Shahabad district, and a simple straight line for the Monghyr material. The graduating equations are :—

$y=3{\cdot}66+0{\cdot}41(x)-0{\cdot}011(x^2)$ .. .. .. for Shahabad ;

and $y=3{\cdot}71+0{\cdot}30(x)$ .. .. .. for Monghyr,

where y is the yield of grain of wheat in maunds per acre, and x the anna-composition of wheat in mixed crops. It may be mentioned here that the value of the coefficient of correlation was +0·269 for the Shahabad material and +0·370 for the Monghyr material ; standard deviations of x (the anna-composition) were 4·22 and 4·18 for the two districts respectively ; and standard deviations of y were 3.39 and 3.14 maunds per acre for the two districts respectively. The graduated values are given in col. (4) for Shahabad and in col. (7) for Monghyr.

195. One point deserves notice in Table (7/3). The rate of yield per acre does not increase in equal proportion with the increase in the anna-composition of cultivation. For example, in Shahabad district, the graduated yield is 4·3 maunds per acre for an anna-composition of 1 and 2-anna proportion of wheat in mixed crops ; 6·1 maunds per acre for an anna-composition 7 and 8-anna ; and 7·4 maunds per acre for an anna-composition of 15 and 16-anna. In the same way the corresponding yields are 4·2, 6·0 and 8.4 maunds per acre in Monghyr district. Thinner compositions evidently give proportionately greater yields which is probably the reason for their popularity.

CALCULATION OF OUTTURN FROM MIXED CROPS

196. From a knowledge of the way in which the rate of yield per acre depends on the anna-composition of wheat it is possible to calculate the total outturn without any difficulty. We had seen in Table (7/4) that 6·5 per cent of wheat plots in Shahabad had been recorded as having an effective anna-composition of 5-anna or less. The mid-value of the composition in this category is 3-anna which may be adopted as the average composition. The total area of 300·3 thousand acres under wheat in Shahabad and 284·1 thousand acres under wheat in Monghyr can now be allocated to each category in proportion to the respective percentage of sample plots shown in col. (3) of Table (7/4). For example, the allocated area in the category "1-anna to 5-anna" (with mid-value 3-anna) is 21·1 thousand acres. The corresponding gross area must be, of course, much higher ; and can be reconstructed by multiplying the net effective area 21·1 by 16/3 which gives 112·5 thousand acres as shown in col. (5) of Table (7/4). In the same way for the class of 6 to 10-anna (with an average composition of 8-anna) the multiplying factor will be 16/8 ; and for the class of 11 to 15-anna (with an average com- position of 13-anna) the multipllying factor will be 16/13. It is thus possible to pass from the estimated net or effective area to the gross area under wheat.

TABLE (7/4). RECONSTRUCTED VALUES OF GROSS AREA UNDER WHEAT IN THOUSAND ACRES AND OUTTURN OF WHEAT IN THOUSAND MAUNDS

Anna proportion of intensity of cultivation of wheat in mixed crops (sample plots)	Plots under wheat		Area under wheat in thousand acres		Yield in maunds per acre	Estimated outturn in thousand maunds
	Total number	percentage	effective (observed)	gross re-constructed		
1	2	3	4	5	6	7
		DISTRICT SHAHABAD				
1-anna to 5-anna	242	6·5	21·1	112·5	4·79	539
5.3 anna	1847	50·0	150·1	453·1	5·52	2501
6-anna to 10-anna	756	20·4	59·7	119·4	6·24	745
11-anna to 15-anna	530	14·3	42·0	51·7	7·13	369
16-anna	322	8·8	27·4	27·4	7·40	203
Total	3697	100·0	300·3	764·1	5·70	4357
		DISTRICT MONGHYR				
1-anna to 5-anna	235	8·1	22·6	120·5	4·61	556
5·3-anna	1029	35·7	91·1	275·0	5·30	1458
6-anna to 10-anna	989	34·3	89·8	179·6	6·11	1097
11-anna to 15-anna	290	10·1	35·3	43·4	7·61	330
16-anna	341	11·8	45·3	45·3	8·51	386
Total	2884	100·0	284·1	663·8	5·77	3827

197. Results of such calculations are shown in Table (7/4) in which col. (4) gives the estimated net or effective acreage under wheat in different anna classes. Multiplying these by appropriate factors shown in col.(3) we get the gross acreage under wheat shown in col.(5). It will be noticed that in Shahabad district a net or effective acreage of 300·3 thousand acres would be roughly equivalent to a gross area of 764·1 thousand acres. In the same way, a net or effective wheat area of 284·1 thousand acres would be equivalent to a gross area of 663·8 thousand acres in Monghyr district.

198. The method for calculating the total outturn is also shown in Table (7/4) in which col. (6) gives the graduated value of the yield in each anna-class of composition of wheat. Multiplying these rates of yield by the corresponding reconstructed gross acreage given in col. (5) we get the total estimated production of wheat grain in thousand maunds for each category of plots as shown in col. (7). Adding up the figures in col. (7) we get the reconstructed outturn of 4357 and 3827 thousand maunds for the two districts Shahabad and Monghyr respectively.

NEED OF FURTHER STUDIES

199. These figures are much higher than the two previous figures of outturn, namely, 1947 and 1833 thousand maunds respectively in the two districts which we had obtained in Section 6. The method of calculation adopted in the present section should give more correct results; but, owing to paucity of the data on which these particular calculations are based it is not certain how much weight can be attached to the present figures. Calculating the outturn of plots under mixed crops by multiplying the net or effective acreage by the rate of yield for a "pure" crop (*i.e.* sown by itself without mixture) would lead to under-estimation as plots with thinner compositions of wheat give proportionately higher yields than plots with heavier compositions. Owing to the fact that wheat in Bihar is usually sown in mixture with widely varying compositions it is also clear that lumping together of all plots under mixed crops would lead to wrong results. Calculations given in this Report illustrate the many pitfalls arising from the occurrence of mixed crops in the *rabi* season in Bihar.

200. In view of the difficulties in the interpretation of estimates relating to mixed crops it must be emphasized that no valid comparison is possible at present between sample estimates and official figures. Further experimental studies are required before a satisfactory method can be found for dealing with mixed crops of the widely varying kind which occur in Bihar. The subject is of great importance and carefully designed experiments should be undertaken preferably in collaboration with the Agricultural Department.

SECTION 8. COST OF OPERATIONS

201. I have already explained that one of the chief objects of the exploratory survey in the *rabi* season in 1944 was to collect information which would enable the running cost of the survey being reduced in future by more efficient planning of the sampling technique. Records were, therefore, kept of the amount of time required for enumeration of the grids and other items of work. Certain results based on these records are discussed in the present section.

FIELD BRANCH

202. The total number of workers and their distribution among different sub-blocks have been given in Table (4/1) in Section 4. In order to assess the total volume of work it is not enough to give the number of workers ; it is also necessary to give the number of days on which they had actually worked. Here a distinction has to be made between what may be called 'working-days' and 'pay-days'. By 'working day' is meant a day which is actually spent in some kind or other of field work. On the other hand, a 'pay day' is a day for which salary was paid irrespective of whether any actual field work was done or not. It is clear that, owing to minor illnesses and cessation of work through inclement weather and other unforeseen reasons, the number of 'working days' will be less than the number of 'pay-days'. For cost studies it is, however, convenient to use the 'pay-day' as the unit of measurement.

203. A second point must be explained before going into further details. It is clear that there are different types or grades of working days, namely, the working day of a chief inspector or that of an inspector or that of an investigator. Strictly speaking, the whole of the primary work is actually done only by the field investigators, the duty of the chief inspectors and inspectors being simply to see that the work is done properly. It is therefore necessary to distinguish carefully between the two types of work. In fact the most convenient plan is to look upon the 'working day' of the investigators as the fundamental unit of measurement, and treat the 'working-day' of the inspecting staff as overhead.

204. To recapitulate, in the present report one 'working-day' means one day of actual field work done by an investigator and is the fundamental unit of measurement ; but for cost studies, the 'pay-day' of an investigator, namely, a day for which salary is paid to one investigator is adopted as the basic unit throughout the present report.

205. In converting investgator-days or pay-days into money values it is of course necessary to add the whole of the over-head in the way of pay-days of the inspecting and supervising staff as well as all other expenses incurred for travelling and contingent charges. Such conversion of investigator-days into money values is however of importance only for budget purposes. So far as the technical problem or preparing the design of the survey is concerned it is clear that results in terms of pay-days are fully adequate.

TOTAL VOLUME OF WORK

206. The number of investigators and the number of pay-days (as defined above) spent in different operations of the survey are shown in Table (8/1) for the extensive scheme separately for each sub-block in the two districts Shahabad and Monghyr. Col. (1) gives the serial number of the sub-block ; and col. (2) the number of investigators working in each sub-block. The total number of pay-days spent in area survey work is given in col. (3), and that in crop-cutting work in col. (4) ; the total number of pay-days consumed is shown in col. (5). The number of grids actually enumerated in the area survey is shown in col. (6) ; the number of sample-cuts harvested is shown for wheat in col. (7) and for gram in col. (8) ; and the total number in col. (9).

207. It will be noticed that the total number of investigators employed in the extensive scheme was 73, and that they spent a total number of 4,882 pay-days consisting of 4,565 days of work in area survey, and 317 days of work in the crop-cutting scheme. The total number of grids actually enumerated in the area survey was 14,438 ; and the total number of sample-cuts of crop collected was 3,875 of which 2,378 were of wheat and 1,497 of gram.

208. In the same way, the number of investigators and pay-days spent in the special (intensive) scheme in the block camps and special camps in Shahabad and Monghyr districts, and also in the preliminary work in certain adjoining districts are shown in Table (8/2). It will be remembered that the work done in the intensive scheme was located near a number of camps names of which are given in col. (1). The number of investigators and the total number of pay-days spent are given in cols.(2) and (3) respectively. The number of grids enumerated in duplicate for checking is shown in col. (4) ; and the number of sample-cuts of crop collected in col.(5). The total number of plots surveyed in the complete enumeration of all plots in selected areas is shown in col. (6) ; the number of plots which were surveyed twice in col. (7) ; and the total number of plots surveyed in col. (8).

209. In addition to the pay-days shown in Table (8/2) certain precision experiments on crop-cutting work were also done in the two special camps at Gogri and Bikramgunj. In the intensive scheme the total number of pay-days spent was 1999 ; and in the extensive scheme, as already noted, 4565, so that the total number of pay-days used in the whole survey was 6564.

Table (8/1). Distribution of investigators and pay-days by sub-blocks (extensive scheme).

Serial No. of sub-blocks	No. of investi-gators	Pay-days spent: area survey	Pay-days spent: crop-cutting work	Pay-days spent: total	grids (area survey)	Number surveyed, Sample-cuts (crop-cutting): wheat	gram	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
District : Shahabad								
1	6	428	50	478	1268	296	148	444
2	6	432	33	465	1231	270	248	518
3	6	429	29	458	1359	236	116	352
4	6	345	9	354	1295	64	37	101
5	5	376	35	411	944	348	226	574
6	7	360	21	381	1280	52	56	108
District	36	2370	177	2547	7377	1266	831	2097
District : Monghyr								
1	6	424	7	431	1325	24	20	44
2	7	378	34	412	1433	324	215	539
3	6	441	26	467	1368	180	179	359
4	6	331	27	358	970	268	117	385
5	6	298	22	320	1002	148	115	263
6	6	323	24	347	963	168	20	188
District	37	2195	140	2335	7061	1112	666	1778
Two Districts	73	4565	317	4882	14438	2378	1497	3875

Table (8/2). Distribution of investigators and of pay-days by camps (intensive scheme)

Name of camp	Number of investi-gators	Number of pay-days	Area survey : no. of grids	Crop-cutting no. of cuts	Number of plots in complete enumeration: single survey	duplicate survey	effective total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Block camp parties							
Jamui	4	119	—	786	6009	1855	7864
Lakhisarai	3	100	186	—	3792	772	4564
Khagaria	3	53	—	—	6575	—	6575
Shahpur	3	219	—	2040	5871	1608	7479
Mohania	4	220	318	914	2781	860	3641
Sasaram	4	217	—	—	3254	1073	4327
Total	21	928		3740	28282	6168	34450
Special camp parties							
Gogri	7	401		1380+957*	27654	50	27704
Bikramganj	7	267		736*	7177	—	7177
Total	14	668		1380+1693*	34831	50	34881
Moving camp parties							
Hilsa, Bettiah, Jahanabad	7	403	—	—	28,517	—	38,517
Total	42	1999	—	5120+1693*	91,630	6218	97,848

* multiple cuts of different sizes.

DISTRIBUTION OF PAY-DAYS BY SUB-BLOCKS AND HALTAGES

210. Having considered the total volume of work we may now look to the distribution of pay-days by sub-samples and sub-blocks in the two districts Shahabad and Monghyr which is shown in Table (8/3). Col.(1) gives the serial number of the sub-block. For sample (A) the total number of grids enumerated is shown in col. (2), and the number of pay-days consumed in col. (3). Dividing the number of pay-days by the number of grids surveyed is obtained the number of pay-days consumed per grid which is shown in col. (4). In the same way, the number of grids surveyed, the number of pay-days consumed, and the number of pay-days per grid are shown in cols. (5), (6) and (7) respectively for sample (B). Finally, similar data are given for the two samples (A) and (B) combined in cols. (8), (9) and (10) respectively.

It will be noticed that the number of pay-days consumed per grid, that is, the average number of grids enumerated was nearly the same in each of the sub-blocks in both districts.

DISTRIBUTION OF GRIDS AND PAY-DAYS BY SUB-SAMPLES AND HALTAGES

211. The distribution of grids and pay-days is shown by sub-samples and haltages in Table (8/4) in which col. (1) gives the serial number of haltages instead of the serial number of sub-blocks as in Table (8/3); otherwise the arrangement is identical with that in Table (8/3). Thus for sample (A) the total number of grids surveyed, the number of pay-days consumed, and pay-days consumed per grid are given as before in cols. (2), (3) and (4) respectively. Corresponding figures for exsample (B) are given in cols. (5), (6) and (7); and for the two samples (A) and (B) combined together in cols. (8), (9) and (10) respectively.

TABLE (8/3). DISTRIBUTION OF GRIDS AND PAY-DAYS BY SUB-SAMPLES AND SUB-BLOCKS

Serial No. of sub-blocks	Sample (A)			Sample (B)			Combined Samples (A+B)		
	Number of		pay days per grid	Number of		pay days per grid	Number of		pay days per grid
	grids	pay days		grids	pay days		grids	pay days	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				DISTRICT : SHAHABAD					
1	628	210	·33	640	218	·34	1268	428	·34
2	647	230	·36	584	202	·35	1231	432	·35
3	624	214	·34	735	215	·29	1359	429	·32
4	808	207	·26	487	138	·28	1295	345	·27
5	392	161	·41	552	215	·39	944	376	·40
6	762	205	·27	518	155	·30	1280	360	·28
District	3861	1227	·32	3516	1143	·33	7377	2370	·32
				DISTRICT : MONGHYR					
1	517	208	·40	808	216	·27	1325	424	·32
2	745	200	·27	688	178	·26	1433	378	·26
3	765	225	·29	603	216	·36	1368	441	·32
4	566	172	·30	404	159	·39	970	331	·34
5	597	176	·29	405	122	·30	1002	298	·30
6	544	173	·32	419	150	·36	963	323	·34
District	3734	1154	·31	3327	1041	·31	7061	2195	·31

212. The only point to be noted in this table is the fact that the number of pay-days consumed per grid was appreciably higher (or the number of grids surveyed per day was appreciably lower) in the fifth haltage in each of the samples (A) and (B) in both districts. This is also true of the fourth haltage but to a smaller extent. This is due to the fact that in the fourth and to a greater extent in the fifth haltages the investigators were also required to do crop-cutting work in addition to area enumeration. This naturally slowed down the output.

TABLE (8/4). DISTRIBUTION OF GRIDS AND PAY-DAYS BY SUB-SAMPLES AND HALTAGES.

Serial No. of haltages	Sample (A)			Sample (B)			Combined Samples (A+B)		
	Number of		Pay days per grid	Number of		Pay days per grids	Number of		Pay days per grid
	grids	pay days		grids	pay-days		grids	pay-days	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				DISTRICT : SHAHABAD					
1	854	263	·31	753	243	·32	1607	506	·31
2	816	235	·29	820	227	·28	1636	462	·28
3	883	244	·28	713	229	·32	1596	473	·30
4	781	260	·33	704	232	·33	1485	492	·33
5	527	225	·43	526	212	·40	1053	437	·42
Total	3861	1227	·32	3516	1143	·33	7377	2370	·32
				DISTRICT : MONGHYR					
1	795	242	·30	613	182	·30	1408	424	·30
2	805	246	·31	734	207	·28	1539	453	·29
3	778	222	·29	635	201	·32	1413	423	·30
4	805	244	·30	715	238	·33	1520	482	·32
5	551	200	·36	630	213	·34	1181	413	·35
Total	3734	1154	·31	3327	1041	·31	7061	2195	·31

RELATION BETWEEN COST AND DENSITY OF GRIDS

213. It has been explained in previous sections that a very simple design was intentionally used for the present survey, and no provision was made for detailed studies of the cost of operations for different patterns of field organisation. The over-all density of grids was kept more or less the same, that is, an equal number of sample units was allotted to each zone-cell of equal size. Village maps were, however, not available in certain instances and also for other unforeseen reasons some of the grids had to be abandoned. The number of grids actually enumerated was thus reduced in some of the zone-cells. This introduced variations in the effective density, that is, in the actual number of grids enumerated per square mile. Advantage was taken of such gaps to determine approximately how the cost of operations varied with changes in the density of sample-units.

214. Material relating to the time consumed for different effective densities of grids is shown in Table (8/5). For this purpose the different zone-cells were grouped in accordance with the effective density, that is, the number of grids actually surveyed per square mile. The class range for such grouping is shown in col. (1); and the midpoint of the range in col. (2). Thus the ranges for classification were effective densities of 0·21 to 0·40, 0·41 to 0·60, . . 2·41 to 2·60 with respective midpoints 0·305, 0·505, . . 2·505. The respective number of zone-cells in each class in district Shahabad is shown in col. (3) from which we find, for example, that the number of zone-cells having effective densities of grids actually enumerated between 1.41 and 1·60 (with a mid-value of 1·505) was 12; and so on. The average number of pay-days consumed per grid for each class of effective density in district Shahabad is shown in col. (4). In the same way, the number of zone-cells in each category and the corresponding number of pay-days consumed per grid for district Monghyr are given in cols. (6) and (7) respectively; and similar data for both districts combined in cols. (9) and (10) respectively.

215. It will be noticed from col. (10) of Table (8/5) that for an effective density of 0·705 the time consumed was 0·34 pay-day per grid, while it is appreciably lower and only 0·27 pay-day per grid for an effective density of 2·505 grids per square mile. Speaking generally, the number of pay-day per grid decreased as the density of grids increased. This is just what one would expect. When the density of grids is high, that is, when there are a large number of grids per square mile, it is obvious that the time spent in moving from one grid to another would be less per grid than when the density is lower and the grids are more widely scattered.

216. The observed values of pay-day per grid were then graduated by a straight line. The calculated parameters are given in Table (8/6). The total number of zone-cells used for purposes of graduation was 96 for Shahabad district. The average cost in terms of pay-day per grid was 0·3142, and the average density of grids per square mile was 1·4925. For Shahabad the regression equation was

$$C = 0{\cdot}3811 - 0{\cdot}0515\,(D) \;\ldots\; (1)$$

where (C) is number of pay-days per grid, and (D) the effective density or number of grids surveyed per square mile. The ratio of variance due to linear regression was 17·83 indicating that the observed decrease in cost (pay-day per grid) with increasing density of grids was statistically significant and must be considered to be real.

217. The graduated values of cost (pay-day per grid) for Shahabad district is shown in col. (5) of Table (8/5). Considering the nature of the material the graduation is not unsatisfactory.

218. The graduation parameters for Monghyr and for the two districts taken together are also given in Table (8/6). For Monghyr also the linear regression is statistically significant; and this is true for the two districts taken together. The two regression equations are :—

$$C = 0{\cdot}3714 - 0{\cdot}0422\,(D) \;\ldots\;\ldots\; (2) \text{ for Monghyr; and}$$
$$C = 0{\cdot}3788 - 0{\cdot}0465\,(D) \;\ldots\;\ldots\; (3)$$

for the two districts taken together, where (C) as before is the cost in pay-day per grid and (D) the effective density of grids surveyed.

219. Graduated values of cost (pay-day per grid) are shown in col. (8) for Monghyr district, and for both districts taken together in col. (11) of Table (8/5). The agreement between observed and graduated values is on the whole satisfactory.

220. Previous work in Bengal had shown that the cost of operations increases with decreasing density, that is, with increasing scatter of the grids. This has been also found to be true in Bihar in the survey in the *rabi* season. The relation between cost and density in Bengal was not however strictly linear, but within the limited range of present observations in Bihar a straight line fit appears to be good enough for all practical purposes.

TABLE (8/5). OBSERVED AND GRADUATED VALUES OF COST (PAY-DAYS PER GRID) FOR DIFFERENT DENSITIES OF GRIDS

Density or number of grids surveyed per square mile		District : Shahabad			District : Monghyr			Both Districts		
		No. of Zone-cells	Pay-days per grid		No. of Zone-cells	Pay-days per grid		No. of Zone-cells	Pay-days per grid	
Range	Mid-point		Observed	Graduated		Observed	Graduated		Observed	Graduated
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
0·21-0·40	0·305	3	0·3667	0·3651	1	·4400	0·3583	4	·3850	0·3644
0·41-0·60	0·505	—	—	·3548	—	—	·3498	—	—	·3551
0·61-0·80	0·705	4	0·3375	·3445	5	·3500	·3414	9	0·3444	·3458
0·81-1·00	0·905	10	·3220	·3342	15	·3173	·3329	25	·3192	·3365
1·01-1·20	1·105	9	·3544	·3239	8	·3200	·3244	17	·3382	·3271
1·21-1·40	1·305	10	·3820	·3136	7	·3014	·3160	17	·3488	·3178
1·41-1·60	1·505	12	·3317	·3033	10	·3480	·3075	22	·3354	·3085
1·61-1·80	1·705	23	·2713	·2930	7	·3389	·2991	30	·2857	·2992
1·81-2·00	1·905	17	·2924	·2827	10	·2650	·2706	27	·2822	·2899
2·01-2·20	2·105	4	·3070	·2724	17	·2741	·2822	21	·2792	·2806
2·21-2·40	2·305	2	·3060	·2621	5	·2720	·2737	7	·2713	·2714
2·41-2·60	2·505	2	·3151	·2518	9	·2678	·2652	11	·2737	·2620

TABLE (8/6). CALCULATED PARAMETERS FOR GRADUATION OF COST (PAY-DAYS PER GRID) ON DENSITY OF GRIDS.

Name of parameter	District : Shahabad	District : Monghyr	Both districts combined
(1)	(2)	(3)	(4)
Total number of Zone-cells	96	94	190
Mean cost in pay-day per grid (C)	0·3142±0·0060	0·3033±0·0051	0·3088±0·0040
Mean density (or number) of grids per sq. mile (D)	1·4925±0·0482	1·6092±0·0559	1·5503±0·0096
Coefficient of correlation between cost (C) and Density (D)	−0·4107	−0·4938	−0·4505
Standard deviation of cost (Sc) and of Density (SD)	=0·0592 ; SD=0·4720	Sc=0·0498 ; Sd=0·5820	Sc=0·0549 ; SD=0·5320
Ratio & Variance for test of linear regression	17·8275	29·6510	48·1566
Regression of cost (c) on Density (D)	C= ·3811− ·0515(D)	C= ·3714− ·0422(D)	C =3788− ·0465(D)

PRACTICAL VALUE OF RESULTS

221. It is worth while pointing out the importance and practical value of the above study. In preparing the design for the *bhadoi* season it was possible to calculate in advance (with the help of the above results) the time which would be required for an area survey with different densities of grids. Information relating to cost of operations collected in the *rabi* season was thus of immediate and direct practical use in preparing the programme of work in the *bhadoi* season.

EXPENDITURE IN THE FIELD BRANCH

222. *Worker-months.* We may now consider briefly the expenditure in money incurred in the field branch for the survey in the *rabi* season of 1943-44. This is shown in Table (8/7) in which col. (1) gives the heads of broad groups of expenditure. Col. (2) shows the number of months of work put in by different categories of workers ; from this column we find that the total number of months of work done by investigators was 158·3, that done by inspectors was 37·0 months, and that by chief inspectors 13·4 months in all. The corresponding volume of work done by each category of field staff in the Bengal Crop Survey of 1943-44 (in two crop seasons, namely, *bhadoi* or jute and *aus* paddy, and *aman* paddy) is shown forcomparison in col. (3) from which we find that the total number of months of work done by investigators was 2000 in round numbers, that done by inspectors 430 months, and by chief inspectors 103 months. The next two columns give the proportion of inspecting staff per investigator-month ; from cols. (4) and (5) we notice that in Bihar the effective proportion of inspectors was 0·24 per investigator-month, that is, one inspector per 4 investigators against 0·22 or some thing between 1 in 4 and 1 in 5 in Bengal. The proportion of chief inspectors in Bihar was 0·09 or 1 chief inspector per 11 investigators against 0·05 or 1 in 20 in Bengal.

TABLE (8/7). COMPARATIVE FIELD EXPENDITURE IN BENGAL AND BIHAR.

Head of expenditure	Number of worker-months		Rate per investigator month		*Total expenditure in Rupees		Rate in rupees per investigator month	
	Bihar *rabi*	Bengal 1943-44	Bihar *rabi*	Bengal 1943-44	Bihar *rabi* 1943-44	Bengal 1943-44	Bihar *rabi* 1943-44	Bengal 1943-44
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Investigators	155·3	2000	1·00	1·00	6,270	72,000	40·50	36·00
2. Inspectors	37·0	430	0·24	0·22	2,330	30,100	15·00	15·05
3. Chief Inspectors	13·4	108	0·09	0·05	998	18,900	6·30	9·45
4. Sub-total (1 to 3)	205·7	2,538	1·33	1·27	9,598	121,000	61·80	60·50
5. Dearness allowance					2,092	34,000	13·50	17·00
6. Leave salary & Provident fund				—	—	6,400	—	3·20
7. Sub-total (5+6)					2,092	40,400	13·50	20·20
8. Field staff sub-total (4+7)					11,690	161,400	75·30	80·70
9. Supervising staff					5,603	6,500	36·00	3·25
10. Headquarters					4,183	7,100	27·00	3·55
11. Sub-total (9+10)					9,786	13,600	63·00	6·80
12. Staff sub-total (8+11)					21,476	175,000	138·30	87·50
Other expenses								
13. Travelling allowance					2,798	20,000	18·00	10·00
14. House rent					332	2,400	2·10	1·20
15. Contingent (with peon)					6,270	25,800	40·50	12·90
16. Other expenses sub-total					9,400	48,200	60·60	24·10
17. Grand Total					30,876	223,200	198·90	111·60

* excludes non-recurring expenditure on training.

223. *Total expenditure.* The total expenditure incurred in Bihar, correct to the nearest rupee, is shown in col. (6); and the corresponding expenditure incurred in the Bengal scheme, in round figures in col. (7). In Bihar there was also an amount of about Rs. 9,052 spent on account of pay and travelling allowances of the field staff sent on loan from Bengal;

this has not been included in Table (8/7) as it was an extra-ordinary item of expenditure incurred for the training of the Bihar staff. In the same way one item of about Rs. 10,000 in Bengal spent for giving training to new workers has been omitted.

224. The total expenditure for the full provincial survey in two seasons in Bengal was about 2·23 lakhs of rupees. The expenditure of Rs. 30,876 for one single survey in only two districts in Bihar was naturally higher in comparison owing to the fact that certain preliminary expenses of a non-recurring nature were incurred in the pilot survey, and also because overhead costs were proportionately much higher on account of the small scale of operations.

225. *Cost per investigator-month.* Dividing the total expenditure by the number of investigator-months we get the "cost per investigator-month" in Bihar and Bengal as shown in cols. (8) and (9) respectively of Table (8/7). We notice that in Bihar the cost of salary of investigators was Rs. 40·5 against Rs. 36 per month in Bengal in 1943-44; the proportionate share of the cost of salary of inspectors was Rs. 15 per investigator-month in both provinces; the proportionate cost of the salary of chief inspectors was Rs. 6·3 in Bihar against Rs. 9·45 in Bengal. The total cost on account of salary of field staff was Rs. 61·8 in Bihar against Rs. 60·5 in Bengal per investigator-month. Adding dearness allowance of Rs. 13·5 (which includes the allowance not only for the investigators but also a proportionate share of the allowance for inspectors and chief inspectors) the total expenditure on account of staff salary was Rs. 75·3 per investigator-month in Bihar. The dearness allowance was higher, Rs. 17 in Bengal; there was also an additional item of Rs. 3·2 as contribution to leave salary and provident fund per investigator-month. The total expenditure on account of salary of field staff was thus appreciably higher, Rs. 80·7 per investigator-month in Bengal.

226. There was not much difference in the salary of the supervising staff, Rs. 5,603 in Bihar against R.s 6,500 in Bengal. But the per capita expenditure was naturally very much higher (because of the small number of investigators) in Bihar, Rs. 36 against Rs. 3·25 in Bengal. Big differences naturally again occured on a per capita basis in expenditure incurred for staff at headquarters with Rs. 27 per investigator-month in Bihar against only Rs. 3·55 in Bengal. Among other expenses, travelling allowance consumed Rs. 18 in Bihar against Rs. 10 in Bengal; house rent was Rs. 2·1 in Bihar against Rs. 1·2 in Bengal. Contingent charges (including pay and allowance of peon) were proportionately much higher in Bihar, Rs. 40·5 per investigator-month against Rs. 12·90 in Bengal, again because of the small scale of operations. The total expenditure under heads other than salary was thus much higher, Rs. 60·6 per investigator-month in Bihar against Rs. 24·10 in Bengal. The total cost per investigator-month was Rs. 198·9 in Bihar against Rs. 111·60 in Bengal.

227. *Expenses for crop-cutting work.* It is of interest to note that crop cutting work consumed only 317 pay-days out of a total of 6,700 pay-days in the whole survey. It will be remembered that the precision in estimating crop acreage and average yield per acre for the province as a whole was about the same. The area survey work thus required about 14 times more labour than crop-cutting work to attain the same precision in the final results. Also 317 pay-days mean about 10½ investigator-months. As the overall rate of expenditure was Rs. 198·9 per investigator-month the marginal share of expenditure for crop-cutting work was only a little over Rs. 2,100. Crop-cutting work is thus not at all expensive.

EXPENDITURE IN THE STATISTICAL BRANCH

228. The total amount of money received by the Indian Statistical Institute up to 31st March 1944 on account of the Bihar Crop Survey scheme was Rs. 61,986 out of which Rs. 5,666-11-0 was credited as contribution in lieu of the Statistical Adviser's services.

229. *Cost per computer-month.* In the statistical branch also the real unit of measurement is the computer-month, that is, the work done by one computer in one month. The total volume of statistical work actually consumed for the survey in the *rabi* season was 241 computer-months between practically January and March 1944. This is the net amount of time spent by primary computers; on an average about 50 per cent of additional staff is required for the inspecting and supervising work in the computation section and in office and general administrative work. For the portion of the work done before 31st March 1944 the adjusted rate of expenditure was about Rs. 150 per computer-month made up of Rs. 95·7 for staff salary (inclusive of pay of primary computers, inspectors, and office and other staff) and Rs. 54·3 for other expenses.

230. It will be noticed that the staff salary per computer-month of Rs. 95·7 was somewhat higher than Rs. 87·5 per investigator-month in the field branch in the Bengal scheme; *pro rata* expenditure on heads other than salary was also higher, Rs. 54·3 per computer-month against Rs. 24·1 per investigator-month in the field branch. Expenses

were naturally higher on the statistical side because men with higher qualifications, as also calculating machines, photographic and other more expensive apparatuses were required for the work.

231. In the Bihar *rabi* survey scheme the total volume of work in the statistical branch was larger than the total volume of field work, 241 computer-months against 155·3 investigator-months. The cost per computer-month, Rs. 150, was somewhat lower than the cost per investigator-month, Rs. 198·9 in the field branch. It must be remembered, however, that so far as the Government of Bihar was concerned, there was also the contribution of Rs. 5,666-11-0 in lieu of the services of the Statistical Adviser. If this is taken into consideration, there would be an additional cost of Rs. 23·5 per computer-month bringing up the total cost to Rs. 173·5 per computer-month. Making allowances for the larger volume of the statistical work, the cost per primary worker in the field and the statistical sides of the Bihar crop survey scheme was not very dissimiliar.

232. A good deal of statistical work has also been done in connexion with the *rabi* survey scheme since 1st April 1944, but detailed figures of cost accounting are not yet available. It may be, however, stated broadly that the total volume of work (inclusive of 241 computer-months consumed upto 31st March 1944) has probably exceeded the estimated figure of 300 computer-months given in the budget.

SECTION 9. ACCURACY OF AREA ENUMERATION

233. A good deal of material was collected in the intensive schemes for special studies for the improvement of the technique of the crop survey. Some of the results of such special studies are given in the present section. It may be mentioned, however, that work is still proceeding in certain directions but progress has been and is still slow as a sufficient number of men cannot be put on to this work for lack of accommodation at Giridih.

STUDY OF COMPLETE AREA ENUMERATION

234. Compact areas were completely enumerated plot by plot in the neighbourhood of the special block camps by a number of field parties employed exclusively for this purpose. In this scheme detailed plot by plot records were prepared showing what crop was grown on what proportion of land in each plot and what proportion was fallow or was being used for other purposes. A complete record was thus prepared for each plot showing in detail how it was being utilized.

235. In certain areas the complete enumeration work was duplicated independently by a second party of investigators. In the original programme the area to be surveyed in each camp was to be covered in sectors, one sector each being alloted to one investigator so that all the sectors would be duplicated in turn in a balanced pattern. Unfortunately, owing to unforeseen difficulties, only a fraction of the field programme relating to the second survey could be completed.

COMPARISON BY INDIVIDUAL PLOTS

236. Duplicate records were however obtained for 6,188 plots by different investigators working independently. This material is of considerable interest as it throws much light on the accuracy with which complete enumeration work can be actually carried out in practice by investigators of the present type appointed *ad hoc* for this purpose.

237. Results of a detailed comparison of the two independent enumerations of the same identical set of individual plots are shown in Tables (9/1)—(9/5). In Table (9/1) the records have been tabulated in a two-way form. An actual example will make the procedure clear. Let us consider, for example, table (9/1a) which gives a comparison of the two sets of records for 1,855 plots surveyed in thana Jamui of district Monghyr, once between 12th and 24th February and a second time between 25th February and 6th March. Entries relating to the wheat crop have been analysed in this particular table. The figures 0-anna, 1-anna, 2-anna, 3-anna....16-anna at the top of columns represent the anna-estimate of the proportion of land sown with wheat as recorded in the first enumeration carried out between 12th and 24th February. In the same way the figures 0-anna, 1-anna, 2-anna....16 anna given in the first column of the table represent for each row the anna-estimate of land under wheat as recorded during the second enumeration carried out between 25th February and 6th March 1944. The figures given in the different cells represent actual number of plots recorded to have the anna-proportions under wheat as shown in the marginal anna figures in the top row and in the first column.

238. Thus we find in the left hand corner cell belonging to '0-anna' row and '0-anna' column the figure 1,573. This means that 1,573 plots were recorded as having '0-anna', that is as being without any wheat in both the enumerations. Just below this figure, in the

column headed '0-anna' and in the row marked '1-anna', we have the figure 7 which shows that 7 plots were recorded as having 'no wheat' in the first enumeration but were recorded as having 1-anna proportion of land under wheat in the second enumeration. In the same way we have the figure 126 in the column headed '0-anna' and row marked '16-anna' showing that these 126 plots were recorded as having 'no wheat' during the first enumeration but were shown as having a 16-anna wheat crop in the second enumeration. At the bottom of the first column we have the figure 1,751 giving the total number of plots which were recorded to have 'no wheat crop' in the first enumeration.

239. Let us now follow the first row marked '0-anna' on the extreme left. The first number is 1,573 which, as already noted, gives the number of plots shown to have 'no wheat' in both enumerations. To the right of this we have the number 13 in the column headed '1-anna' (and the row marked '0-anna') giving the number of plots which were recorded as having 1-anna under wheat in the first enumeration, but were entered as having no wheat at all in the second enumeration, and so on. In the '0-anna' row we have in the column headed '16-anna' the figure 18 giving the number of plots which were recorded as having a 16-anna wheat in the first enumeration but were shown as having no wheat at all in the second enumeration. On the extreme right in the '0-anna' row we have the figure 1,637 giving the total number of plots which were recorded as having no wheat at all in the second enumeration.

DISCREPANCIES IN ANNA-ESTIMATE

240. From the way in which table (9/1a) has been prepared it is clear that if the two enumerations were in complete agreement then entries would have occurred only in the diagonal cells, that is, in the 0-0, 1-1, 2-2. . . . 16-16 cells. That is, the number of plots shown to have no wheat would have been the same in both enumerations. The number of plots shown to have 1-anna under wheat, 2-anna under wheat etc. up to 16-anna under wheat would also have been identical in both enumerations. A glance at tables (9/1a)—(9/1f) shows that in actual fact this was not so ; and in an appreciable number of cases plots were recorded to have different proportions of land under wheat in the two enumerations. As already pointed out, among 1,751 plots shown as not having any wheat in the first enumeration, only 1573 were recorded as having no wheat in the second enumeration ; the remaining 178 plots (which were recorded as having no wheat in the first enumeration) were shown to have different proportions of land from 1-anna to 16-anna under wheat in the second enumeration.

241. It is clear therefore that results of complete enumeration carried out by *ad hoc* investigators of the present type are subject to appreciable mistakes in recording. This is fully borne out by the other tables (9/1b)—(9/1f). Considerable variations occur in the anna estimates of the area under the same crop made by different investigators. This would not have been very serious if the anna estimates had differed by a comparatively small amount. For example, the same plot might easily be estimated to have a 4-anna crop of wheat by one investigator and to have a 5-anna or 6-anna crop by a second investigator. The position as revealed by tables (9/1a) to (9/1f) is more serious. We found that plots recorded as having been sown completely with wheat in the first enumeration were shown as having no wheat at all in the second enumeration. Such discrepancies could have only arisen either through a mistake in the identification of the plot or in the identification of the crop, or through the work having been fudged altogether.

DISCREPANCIES IN CROP RECORDS

242. To investigate the matter in greater detail it was considered advisable to tabulate the duplicate records of the same set of individual plots in a different way in Table (9/2). Instead of classifying under anna-estimates of the proportion of land under the same crop, the material is now classified simply according to the particular crop shown to be growing on the plot irrespective of the proportion of land under different crops. Thus in Table (9/2a) the column headings at the top are simply 'wheat', 'gram', 'wheat and gram', 'wheat, gram, and others', 'gram and others', 'other crops', and 'no crop' (the last one signifying that the plots were either kept fallow or were being used for purposes other than cultivation). The same headings also occur in the extreme left against each row.

243. In this Table (9/2) also, if the results of the two enumerations are in complete agreement, then entries would occur only in the diagonal cells, that is, in the cells which are marked both ways as being 'under wheat', or 'under wheat and gram' etc. A glance at Tables (9/2a) to (9/2f) shows that quite a large number of entries occur outside the diagonal cells. We thus find that, apart from discrepancies in anna-estimates, appreciable discrepancies had also occurred in recording the crops grown on the same individual plots enumerated by different sets of investigators.

TABLE (9/1). COMPARISON OF DUPLICATE ANNA ESTIMATES OF PROPORTIONS OF LAND UNDER WHEAT

TABLE (9/1a). THANA JAMUI, DISTRICT MONGHYR

Anna estimates of proportion of land under wheat made in 2nd enumeration from 25 Feb. to 6 March 1944	Anna estimates of proportion of land under wheat as recorded in first enumeration between 12 February and 24 February 1944																	
	0-anna	1-anna	2-anna	3-anna	4-anna	5-anna	6-anna	7-anna	8-anna	9-anna	10-anna	11-anna	12-anna	13-anna	14-anna	15-anna	16-anna	Total
0-anna	[1573]	13	8	1	3	—	3	1	7	—	3	—	5	2	—	—	18	1637
1-anna	7	[—]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
2-anna	9	1	[—]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
3-anna	—	—	—	[—]	—	—	—	—	—	—	—	—	—	—	—	—	—	—
4-anna	2	—	1	—	[—]	—	1	—	—	—	—	—	—	—	—	—	—	4
5-anna	3	—	—	—	1	[—]	—	—	—	—	—	—	—	1	—	—	—	5
6-anna	2	—	—	—	—	—	[—]	—	—	—	—	—	—	—	—	—	—	2
7-anna	1	—	—	—	—	—	—	[—]	1	—	—	—	—	1	—	—	—	3
8-anna	3	—	—	—	2	—	—	—	[1]	—	—	—	2	—	—	—	1	9
9-anna	—	—	—	—	—	—	—	—	—	[—]	—	—	1	—	—	—	1	2
10-anna	5	—	—	—	—	—	1	—	—	—	[—]	—	1	—	—	—	—	7
11-anna	2	—	—	—	—	—	—	—	1	—	—	[—]	—	—	—	—	—	3
12-anna	8	—	—	—	—	—	2	—	1	—	—	—	[1]	—	1	—	1	14
13-anna	1	—	—	—	—	—	—	—	—	—	—	—	—	[—]	1	—	—	2
14-anna	6	—	—	—	—	—	—	1	—	—	—	—	3	1	[—]	—	1	12
15-anna	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	3
16-anna	126	—	1	—	—	—	1	—	4	—	—	—	1	—	—	—	[2]	135
Total	1751	14	10	1	6	—	8	2	15	—	3	—	14	5	2	—	24	1855

TABLE (9/1b). THANA SHAHPUR, DISTRICT SHAHABAD

Anna estimates of proportion of land under wheat made in 2nd enumeration from 23 February to 6 March 1944	Anna estimate of proportion of land under wheat as redcorded in first enumeration between 13 February and 23 February 1944																	
	0-anna	1-anna	2-anna	3-anna	4-anna	5-anna	6-anna	7-anna	8-anna	9-anna	10-anna	11-annaz	12-anna	13-anna	14-anna	15-anna	16-anna	Total anna
0-anna	[1389]	10	11	13	11	3	6	8	4	3	1	—	—	—	—	—	—	1459
1-anna	10	[2]	1	—	1	—	1	—	—	—	—	—	—	—	—	—	—	15
2-anna	21	1	[1]	1	1	—	—	1	—	—	—	—	—	—	—	—	—	26
3-anna	22	1	—	[1]	2	1	—	—	—	—	—	—	—	—	—	—	—	27
4-anna	23	3	3	—	[—]	—	1	1	—	—	1	—	1	—	—	—	—	33
5-anna	8	—	—	1	2	[—]	1	—	—	—	1	—	—	—	—	—	—	13
6-anna	7	2	1	1	—	—	[1]	3	—	—	1	1	—	1	—	—	—	18
7-anna	4	—	—	1	—	—	—	[—]	—	—	—	—	—	—	—	—	—	5
8-anna	4	—	—	—	1	1	—	1	[1]	—	—	—	—	—	—	—	—	8
9-anna	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—	—	—
10-anna	2	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—	2
11-anna	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—
12-anna	—	—	—	—	—	—	—	—	—	—	—	1	[—]	—	—	—	—	1
13-anna	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—
14-anna	1	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—	1
15-anna	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—
16-anna	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—
Total	1491	19	17	18	18	5	10	14	5	3	4	2	1	1	—	—	—	1608

TABLE (9/1c). THANA SASARAM, DISTRICT SHAHABAD

Anna estimates of proportion of land under wheat made in 2nd enumeration from 25 February to 6 March 1944	Anna estimates of porportion of land under wheat as recorded in first enumeration between 11 February and 23 February 1944																	
	0-anna	1-anna	2-anna	2-anna	4-anna	5-anna	6-anna	7-anna	8-anna	9-anna	10-anna	11-anna	12-anna	13-anna	14-anna	15-anna	16-anna	Total
0-anna	[773]	5	4	1	3	—	6	2	8	1	5	1	3	—	5	2	5	824
1-anna	2	[1]	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	5
2-anna	3	1	[—]	—	2	—	1	—	2	—	—	—	—	—	—	—	—	9
3-anna	—	—	—	[—]	1	2	—	1	1	—	—	—	—	—	—	—	—	5
4-anna	1	—	1	1	[—]	—	1	—	—	1	—	—	—	—	—	—	—	5
5-anna	1	—	1	—	2	[1]	1	—	1	2	—	—	1	—	—	—	—	10
6-anna	3	—	1	1	2	2	[1]	1	1	2	1	1	—	—	—	—	—	16
7-anna	1	—	—	—	—	1	2	[—]	—	1	4	—	—	1	—	—	—	10
8-anna	7	—	—	—	—	3	1	1	[3]	3	2	1	—	—	2	3	—	26
9-anna	1	—	1	—	1	—	2	—	5	[—]	3	—	1	—	—	—	—	14
10-anna	2	—	1	—	—	2	—	2	5	3	[2]	3	2	—	1	—	—	23
11-anna	—	—	—	—	1	1	1	—	1	1	9	[2]	1	—	1	—	1	19
12-anna	2	—	—	—	2	4	—	—	7	—	8	2	[3]	1	—	—	—	29
13-anna	1	—	—	—	—	—	1	1	2	4	2	—	6	[3]	2	—	1	23
14-anna	—	—	—	—	—	—	—	—	1	1	1	—	—	—	[1]	—	—	4
15-anna	2	—	—	—	—	—	—	1	—	—	—	—	1	—	1	[—]	4	9
16-anna	8	1	1	—	—	—	1	—	—	—	—	—	—	—	2	7	[22]	42
Total	807	8	11	3	15	16	18	9	37	19	37	10	18	5	15	12	33	1073

TABLE (9/1d). THANA LAKHISARAI, DISTRICT MONGHYR

Anna estimates of proportion of land under wheat made in 2nd enumeration from 25 February to 6 March 1944	Anna estimates of proportion of land under wheat as recorded in first enumeration between 12 February and 24 February 1944																	
	0-anna	1-anna	2-anna	3-anna	4-anna	5-anna	6-anna	7-anna	8-anna	9-anna	10-anna	11-anna	12-anna	13-anna	14-anna	15-anna	16-anna	Total
0-anna	[724]	—	2	—	9	—	1	—	6	—	—	—	1	—	—	—	6	749
1-anna	1	[—]	—	—	1	—	—	—	—	—	—	—	1	—	—	—	—	3
2-anna	1	—	[—]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
3-anna	2	—	—	[—]	—	—	—	—	—	—	—	—	—	—	—	—	—	2
4-anna	—	—	—	—	[1]	—	—	—	—	—	—	—	—	—	—	—	—	1
5-auna	—	—	—	—	1	[—]	—	—	—	—	—	—	—	—	—	—	—	1
6-anna	1	—	—	—	—	—	[—]	—	—	—	—	—	—	—	—	—	1	2
7-anna	—	—	—	—	—	—	—	[—]	1	—	—	—	—	—	—	—	—	1
8-anna	—	—	1	—	—	—	—	—	[1]	—	—	—	—	—	—	—	—	2
9-anna	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—	—	—
10-anna	1	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—	1
11-anna	1	—	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	1
12-anna	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—
13-anna	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—
14-anna	2	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—	2
15-anna	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—
16-anna	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	2
Total	735	—	3	—	12	—	1	—	8	—	—	—	2	—	—	—	7	768

TABLE (9/1e). THANA GOGRI, DISTRICT MONGHYR

Anna estimates of proportion of land under wheat made in 2nd enumeration from 8 Mar. to 20 Mar. 1944	Anna estimates of proportion of land under wheat as recorded in first enumeration between 26 February and 7 March 1944																	
	0-anna	1-anna	2-anna	3-anna	4-anna	5-anna	6-anna	7-anna	8-anna	9-anna	10-anna	11-anna	12-anna	13-anna	14-anna	15-anna	16-anna	Total
0-anna	[27]	—	—	—	—	—	2	1	1	—	—	—	—	1	1	—	3	36
1-anna	—	[—]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2-anna	—	—	[—]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3-anna	1	—	—	[—]	—	—	—	—	—	—	—	—	—	—	—	—	—	1
4-anna	—	—	—	—	[—]	1	1	—	—	—	—	—	—	—	—	—	—	2
5-anna	—	—	—	—	1	[—]	—	—	—	—	—	—	—	—	—	—	—	1
6-anna	—	—	1	—	—	—	[—]	—	—	—	—	—	—	—	—	—	—	1
7-anna	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—	—	—	—	—
8-anna	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—	—	—	—
9-anna	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—	—	—
10-anna	—	—	—	—	1	—	—	—	—	—	[—]	—	—	—	—	—	—	1
11-anna	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—
12-anna	—	—	—	—	—	—	—	—	—	—	—	—	[—]	1	1	—	1	3
13-anna	1	—	—	—	—	—	—	—	—	—	—	—	—	[—]	1	—	—	2
14-anna	1	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	2	3
15-anna	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—
16-anna	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—
Total	30	—	1	—	2	1	3	1	1	—	—	—	—	2	3	—	6	50

TABLE (9/1f). THANA MOHANIA, DISTRICT SHAHABAD

Anna estimates of proportion of land under wheat made in 2nd enumeration from 25 Feb. to 6 Mar. 1944	Anna estimates of proportion of land under wheat as recorded in first enumeration between 13 February and 23 February 1944																	
	0-anna	1-anna	2-anna	3-anna	4-anna	5-anna	6-anna	7-anna	8-anna	9-anna	10-anna	11-anna	12-anna	13-anna	14-anna	15-anna	16-anna	Total
0-anna	[745]	5	3	3	10	1	7	1	9	—	—	—	—	—	1	—	1	786
1-anna	8	[1]	1	3	1	1	—	—	—	—	—	—	—	—	—	—	—	15
2-anna	7	—	[2]	1	1	—	—	—	—	—	—	—	—	—	—	—	—	11
3-anna	4	—	—	[—]	—	2	1	1	1	—	—	—	—	—	—	—	—	9
4-anna	4	—	—	—	[1]	1	—	—	1	—	—	—	1	—	—	—	—	8
5-anna	1	—	—	1	—	[—]	—	—	1	—	1	—	—	—	—	—	—	4
6-anna	1	—	—	—	1	—	[—]	—	—	1	1	—	1	—	—	—	—	5
7-anna	2	—	—	—	—	—	—	[1]	—	—	—	—	—	—	—	—	—	3
8-anna	4	—	—	1	—	—	1	1	[—]	—	—	—	1	—	—	—	—	8
9-anna	—	—	—	—	—	—	—	—	2	[—]	1	—	—	—	—	—	—	3
10-anna	1	—	—	—	—	—	—	—	—	1	[1]	—	—	—	—	—	—	3
11-anna	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—	—
12-anna	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—	—	—	—
13-anna	—	—	—	—	—	—	—	—	1	—	—	—	—	[—]	—	—	1	2
14-anna	—	—	—	—	—	—	—	—	—	—	—	1	1	—	[—]	—	—	2
15-anna	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	[—]	—	—
16-anna	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	[—]	1
Total	777	6	6	9	14	5	9	4	15	2	4	1	5	—	1	—	2	860

TABLE (9/2). COMPARISON OF DUPLICATE CROP RECORDS

Crop records in second enumeration	Crop records in first enumeration							
	Wheat	Gram	Wheat & gram	Wheat, Gram & Others	Gram & others	Other crops	No crops	Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

TABLE (9/2a). POLICE STATION MOHANIA, DISTRICT SHAHABAD
PERIOD OF FIRST ENUMERATION : 13 FEBRUARY TO 23 FEBRUARY 1944

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Period of 2nd enumeraton: 23 Feb. to 6 March 1944	Wheat	[—]	—	—	1	—	1	—	2
	Gram	—	[20]	—	1	6	1	2	30
	Wheat & Gram	1	—	[17]	2	10	5	4	39
	Wheat, Gram & others	—	1	3	[16]	18	2	3	43
	Gram & Others	—	31	2	17	[212]	10	13	285
	Other crops	—	8	1	2	6	[111]	38	166
	No Crop	—	1	1	4	5	8	[277]	296
	Total	1	61	24	43	257	138	377	861

TABLE (9/2b). POLICE STATION SASARAM, DISTRICT SHAHABAD
PERIOD OF FIRST ENUMERATION : 11 FEBRUARY TO 23 FEBRUARY 1944

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Period of 2nd enumeration : 25 Feb. to 6 March 1944	Wheat	[22]	—	—	2	1	13	4	42
	Gram	—	[—]	—	—	1	—	1	2
	Wheat & Gram	1	1	[1]	3	—	2	—	8
	Wheat, gram & others	1	—	23	[98]	1	23	5	151
	Gram & Others	—	—	—	12	[8]	5	1	26
	Other crops	6	—	3	15	3	[145]	39	211
	No crop	3	—	—	6	2	57	[565]	633
	Total	33	1	27	136	16	245	615	1073

TABLE (9/2c). POLICE STATION SHAHPUR, DISTRICT SHAHABAD
PERIOD OF FIRST ENUMERATION : 11 FEBRUARY TO 23 FEBRUARY 1944

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Period of 2nd enumeration: 25 Feb. to 6 March 1944	Wheat	[—]	—	—	—	—	—	—	
	Gram	—	[8]	—	—	12	—	1	21
	Wheat & Gram	—	—	[—]	—	—	—	1	1
	Wheat, Gram & others	—	1	—	[38]	68	22	2	131
	Gram & others	—	7	—	13	[184]	122	25	351
	Other crops	—	1	—	4	39	[240]	31	315
	No crop	—	4	—	3	12	93	[667]	779
	Total	—	21	—	58	315	477	727	1598

244. The two-way tables discussed above show the details of the comparison of duplicated records for individual plots. A more summary form of the tȧbulation is, however, convenient for a general appreciation of the position. This is shown in Table (9/3a) in which the results of the comparison between records obtained during the first and the second enumerations are shown in a condensed form for each camp or centre separately and also for the the districts separately, and finally for the two districts taken together. In this Table col. (1) gives the name of the camp, and cols. (2) and (3) the dates of the first and the second enumeration, respectively. Col. (4) gives the number of 0-0 entries, that is, the number of plots which were recorded as having no wheat in both enumerations. In col. (5) is given the number of plots for which anna-estimates of the proportion of land under wheat was the same in both the enumerations. For these plots the two enumerations were therefore in complete agreement. The number of plots for which the records of the first and second enumerations were not in agreement is shown in col. (6). The total number of plots compared is given in col. (7) of Table (9/3a).

245. Instead of actual numbers, for purpose of comparison, it is more convenient to use percentage figures which are shown in cols. (8), (9) and (10). Col. (8) gives the percentage of 0-0 plots, col. (9) the percentage of plots recorded as having wheat in both enumerations and finally col. (10) of Table (9/3a) shows the percentage of plots for which records in the two enumerations were not in agreement.

TABLE (9/2). COMPARISON OF DUPLICATE CROP RECORDS(contd.)

TABLE (9/2d). THANA JAMUI DISTRICT MONGHYR
PERIOD OF FIRST ENUMERATION : 12 FEBRUARY TO 24 FEBRUARY 1944.

	Crop records in second enumeration	Crop Records in first enumeration							
		Wheat	Gram	Wheat & Gram	Wheat, Gram & Others	Gram & others	Other Crops	No Crops	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Period of 2nd enumeration: 25th Feb to 8th Mar. 1944	Wheat	[4]	—	6	1	1	50	47	109
	Gram	1	[24]	4	—	1	16	20	66
	Wheat & Gram	5	—	[12]	4	3	3	6	33
	Wheat, gram & others	—	—	4	[1]	4	3	12	24
	Gram & others	4	1	4	5	[14]	6	27	61
	Other crop	6	3	5	9	17	[28]	108	276
	No crop	4	9	7	9	26	77	[1154]	1286
	Total	24	37	42	29	66	283	1374	1855

TABLE (9/2e). THANA LAKHISARAI DISTRICT MONGHYR
PERIOD OF FIRST ENUMERATION : 12 FEBRUARY TO 24 FEBRUARY 1944

	Crop records in second enumeration	Wheat	Gram	Wheat & Gram	Wheat, Gram & Others	Gram & others	Other Crops	No Crops	Total
Period of 2nd enumeration : 25thFeb to 6th Mar. 1944	Wheat	[—]	—	3	—	—	—	2	5
	Gram	—	[—]	—	—	—	—	1	1
	Wheat & gram	—	1	[5]	—	—	—	2	8
	Wheat, gram & others	2	—	—	[—]	—	—	3	5
	Gram & others	1	—	5	2	[—]	4	14	26
	Other crops	2	—	3	1	1	[2]	12	21
	No Crop	2	3	5	—	1	3	[692]	706
	Total	7	4	21	3	2	9	726	772

TABLE (9/2f). THANA GOGRI : DISTRICT MONGHYR
PERIOD OF FIRST ENUMERATION : 26 FEBRUARY TO 7 MARCH 1944.

	Crop records in second enumeration	Wheat	Gram	Wheat & Gram	Wheat, Gram & Others	Gram & others	Other Crops	No Crops	Total
Period of 2nd enumeration: 8th Mar. to 20th Mar. 1944	Wheat	[—]	—	—	—	—	—	—	—
	Gram	—	[—]	—	—	—	—	—	—
	Wheat & Gram	2	—	[3]	—	—	—	2	7
	Wheat, Gram & others	1	—	—	[1]	—	4	1	7
	Gram & others	—	—	—	—	[—]	—	—	—
	Other crops	—	—	—	—	—	[—]	—	—
	No crops	3	—	2	2	—	4	[25]	36
	Total	6	—	5	3	—	8	28	50

246. Looking at col. (10) of Table (9/3a) it will be noticed that the proportion of plots not in agreement varied from 6 per cent in Lakhisarai to 46 per cent in Gogri in Monghyr district. The over-all proportion of discrepant plots was 16·5 per cent in Shahabad and 13·0 per cent in Monghyr with a combined proportion of 15·0 per cent for the two districts taken together. Records of the two complete enumerations were thus discrepant in something like one-sixth of the total number of plots surveyed. This is by no means negligible.

247. It will be remembered that in preparing Table (9/3a) agreement was interpreted in the sense that the actual anna estimate of the proportion of land under wheat should be identical in both the enumerations. This is, however, a very stringent definition. For example, a plot recorded as having 5-anna under wheat in one enumertion and 6-anna under wheat in the other enumeration, would be, according to this definition, considered as discrepant. It is possible to adopt other definitions of agreement. For example, a plot recorded as having wheat in it in both enumerations may be considered to be in agreement irrespective of the fact whether the respective anna-estimates were in agreement or had differed widely. Thus, according to this definition, a plot recorded as having one anna under wheat in one enumeration and as 16-anna under wheat in the other enumeration would

be considered to be in agreement. This is the most liberal interpretation of agreement which one could possibly adopt. Results of the comparison on this basis are shown in Table (9/3b) in which the arrangement of the different columns is exactly similar to that in Table (9/3a); col.(10) as before, gives the percentage of plots for which the records were not in agreement, which, in the present case, means the percentage of plots which were recorded as having wheat in one enumeration but not having wheat at all in the other enumeration. It will be noticed that the proportion of discrepancy defined even in this most liberal way was so high as 10 per cent.

TABLE (9/3a). PROPORTION OF PLOTS IN EXACT AGREEMENT IN ANNA ESTIMATE

Name of Camp	Period of survey		Number of Plots			Total No. of plots compared	Percentage of plots for which both records are		
							in agreement		not in agreement
	first survey	second survey	Zero-Zero (no wheat in both surveys)	agreement in anna estimate	discrepancy in anna estimate		Zero-zero	with wheat	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				District : Shahabad					
Mohania	13/2—23/2	25/2—6/3	745	6	109	860	86·6	0·7	12·7
Shahpur	11/2—23/2	25/2—6/3	1389	6	213	1608	86·4	0·4	13·2
Sasaram	11/1—23/2	25/2—6/3	773	39	261	1073	72·0	3·6	24·4
Total	11/2—23/2	25/2—6/3	2907	51	583	3541	82·1	1·4	16·5
				District : Monghyr					
Lakhisarai	12/2—24/2	25/2— 6/3	724	2	46	772	93·7	0·3	6·0
Gogri	26/2— 7/3	8/3—20/3	27	0	23	50	54·0	0·0	46·0
Jamui	12/2—24/2	25/2— 6/3	1573	4	278	1855	48·7	0·2	15·1
Total	12/2— 7/3	25/2—20/3	2324	6	347	2677	86·8	0·2	13·0
Two Districts	11/2— 7/3	25/2—20/3	52231	57	930	6218	84·1	0·9	15·0

TABLE (9/3b). PROPORTION OF PLOTS IN AGREEMENT AS TO OCCURRENCE OF WHEAT

Name of Camp	Period of survey		Number of plots			Total No. of plots compared	Percentage of plots for which both records are		
				having wheat in			in agreement		not in agreement
	first survey	second survey	Zero-zero (no wheat in both surveys)	both surveys	only one survey		Zero-zero	With wheat	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				District : Shahabad					
Mohania	13/2—23/2	25/2—6/3	745	42	73	860	86·0	4·9	8·5
Shahpur	11/2—23/2	25/2—6/3	1389	47	172	1608	84·4	2·9	10·7
Sasaram	11/2—23/2	25/2—6/3	773	215	85	1073	72·1	20·0	7·9
Total	11/2—23/2	25/2—6/3	2907	304	330	3541	82·1	8·6	9·3
				District : Monghyr					
Lakhisarai	12/2—24/2	25/2— 6/3	724	12	36	772	93·7	1·6	4·7
Gogri	26/2— 7/3	8/3—20/3	27	11	12	50	54·0	22·0	24·0
Jamui	12/2— 7/3	25/2— 6/3	1573	40	242	1855	84·70	2·2	13·1
Total	12/2— 7/3	25/2—20/3	2324	63	290	2677	86·8	2·8	10·8
Two districts	11/2— 7/3	25/2— 6/3	5231	367	620	6218	84·1	5·9	10·0

248. Other definitions of agreement intermediate between the two used above are also possible. For example, plots for which the anna estimate of proportion under wheat agreed within 4-anna may be considered to be in agreement ; or the proportion of plots the anna-estimates of which in the two enumerations agreed within 8-anna may be considered as being in agreement, and so on. Results of comparison on the basis of two such definitions, with latitudes of 4-anna and 8-anna respectively, are shown in summary form in Table (9/4). The arrangement in this table is broadly similar to that in Table (9/3). Col. (1) now gives the latitude or margin of agreement. The percentage of plots for which records were discrepant (according to different definitions) are shown in col. (8) of table (9/4).

Table (9/4). Summary of comparison of duplicate records of plots under wheat

Latitude of comparison	Number of plots: in agreement: Zero-zero	Number of plots: in agreement: with wheat	Number of plots: not in agreement	Total No. of plots compared	Percentage of plots: in agreement: Zero-zero	Percentage of plots: in agreement: with wheat	Percentage of plots: not in agreement
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1-anna	5231	57	930	6218	84·1	0·9	15·0
4-anna	5231	164	823	6218	84·1	2·6	13·3
8-anna	5231	239	748	6218	84·1	3·9	12·0
16-anna	5231	367	620	6218	84·1	5·9	10·0

249. For facility of comparison, results of the comparison with the most stringent definition (that is, agreement within 1-anna) and the most liberal definition (agreement only in the presence or absence of wheat, that is, with a latitude of 16-anna) are also shown in the same table (9/4). We notice that with the most stringent definition (requiring complete agreement in the anna-estimate) the over-all proportion of discrepant plots was 15 per cent. Allowing a latitude of 4-anna, this is reduced to 13·3 per cent ; and allowing a latitude of 8-anna, to 12 per cent. Using the most liberal interpretation (namely, either presence or absence of wheat irrespective of the estimated proportion of land under the crop) the number of discrepant plots is 10 per cent which is still not negligible. We find then that whatever definition is adopted the results of two *ad hoc* complete plot-by-plot enumerations had differed in the case of wheat from 10 to 15 per cent of the total number of plots surveyed.

COMPARISON IN TERMS OF ACREAGE

250. So far the records of the two enumerations have been compared in terms of individual plots. A second procedure is also open, namely, a comparison of the total area under wheat (or under any other crop) in any given area as recorded in the two enumerations. Results on this acreage basis of comparison are shown in table (9/5) in which col. (1) gives the name of the police stations, and col. (2) the Jurisdiction List serial number of the village. Col. (3) gives the area in acres under wheat as recorded by the first worker (symbolically represented by W_1) ; and col. (4) the corresponding area entered under wheat in the same village as recorded by the second investigator (W_2). Thus we find from the first line of the table that in police station Sasaram, J. L. No. 121, the area under wheat was recorded as 3.90 acres by the first worker, but the second worker did not find any wheat at all in this particular *mauza*. In the same way, in the second police station in J. L. No. 122/1 the area under wheat was recorded as 2·91 acres by the first and 8·64 acres by the second worker respectively.

251. The actual magnitude of the discrepancy is shown in the next two columns (5) and (6) of Table (9/5). Any area which is recorded to have wheat in the first enumeration but not in the second may be called a positive discrepancy. On the other hand, any area recorded as having no wheat in the first enumeration but as having wheat in the second enumeration may be called a negative discrepancy. The total number of acres with positive and negative discrepancies (defined in the way explained above) are shown in cols. (5) and (6) respectively of Table (9/5). For example, in Sasaram J. L. No. 121 the positive discrepancy is 3·90 acres. In the same way in Sasaram J. L. No. 122/1 the positive discrepancy is 0.87 acre which means that this amount of land was recorded as being under wheat in the first enumeration but not having any wheat in the second enumeration. The negative discrepancy in the same case was 6·60 acres which was the area recorded as being under wheat in the second enumeration but was recorded as having no wheat in the first enumeration.

252. The next two cols. (7) and (8) of Table (9/5) give the sum of discrepancies in two different ways. If we add positive and negative discrepancies retaining their algebraic signs we get the algebraic sum shown in col. (7). In this procedure we allow the positive and negative errors to cancel out as far as possible. In Sasaram J. L. No. 122/1 we had a negative discrepancy of –6.60 acres , adding the positive discrepancy of 0·87 acre we get an algebraic sum of –5·73 acres which is the net discrepancy in the final result. A second procedure is to add the positive and negative discrepancies ignoring their signs. In Sasaram J. L. No. 122/1 we then get 6·60+0·87=7·47 acres shown in col. (8).

TABLE (9/5). COMPARISON OF CROP RECORDS IN TERMS OF ACREAGE : WHEAT

Name of Police Station	J. L. No	Area in acres estimated by		Actual discrepencies (difference between two records)		Sum of discrepancies		Discrepancies as percentage of estimate by first worker	
		first worker (w_1)	second worker (w_2)	positive	negative	algebraic (counting sign of difference	accumulated (neglecting sign of difference)	algebraic	accumulated
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mohania	134	18·10	10·22	9·96	2·08	7·88	12·04	43·54	66·52
	535/1	55·20	61·94	8·17	14·91	–6·74	23·08	–12·21	41·81
	535/2	22·17	18·64	9·22	5·69	3·53	14·91	15·92	67·25
	537	1·78	1·09	0·79	0·10	0·89	0·89	38·76	50·00
Total		97·25	91·89	28·14	22·78	5·36	50·92	5·51	52·36
Shahpur	146	—	7·26	—	7·28	– 7·28	7·28	–100·00	100·00
	147/2	7·86	4·50	4·97	1·61	3·36	6·58	44·55	86·36
Total		7·86	11·78	4·97	8·89	– 3·92	13·86	–50·00	180·91
Sasaram	121	3·90	—	3·90	—	3·90	3·90	100·00	100·00
	122/1	2·91	8·64	0·87	6·60	– 5·73	7·47	–196·91	256·70
	123	2·78	—	2·78	—	2·78	2·78	100·00	100·00
	125	17·98	16·22	3·83	2·07	1·76	5·90	9·79	32·81
	139	21·68	11·31	11·27	0·90	10·37	12·17	47·83	36·13
	140	9·26	13·64	1·65	6·03	– 4·38	7·68	–47·30	82·94
	141/1	74·22	94·17	12·30	32·25	–19·95	44·55	–26·88	60·02
Total		132·73	143·98	36·60	47·85	–11·25	84·45	– 8·48	63·63
Lakhisarai	125/1	6·39	6·59	4·60	4·80	– 0·20	9·40	–3·13	147·10
Gogri	325/1	13·79	11·66	6·51	4·38	2·13	10·89	15·45	78·97
Jamui	10/1	4·75	7·53	1·87	4·65	– 2·78	6·52	–58·53	137·26
	10/2	16·84	6·32	12·07	1·55	10·52	13·62	62·47	80·80
	20	5·43	3·86	4·86	3·29	1·57	8·15	28·91	150·09
	22/2	0·83	14·60	0·83	14·60	–13·77	15·43	–1651·07	1859·03
Total		27·85	32·31	19·63	24·09	– 4·46	43·72	– 16·01	150·99
All Police Stations		285·87	298·21	100·45	112·79	12·34	213·24	4·32	74·59

253. For convenience of comparison these two discrepancies are converted into percentages of the area under wheat as recorded in the first enumeration as given in col. (3) of Table (9/5). The percentage algebraic discrepancy is shown in col. (9) ; and percentage absolute or accumulated discrepancy in col. (10). The algebraic percentage discrepancy is naturally smaller because the positive and negative errors are allowed to cancel out in this case. But, for individual *mauzas* even this algebraic discrepancy can be quite high. For police stations , some of the positive and negative errors for *mauza* as a whole again have the

chance of cancelling out and the percentage algebraic discrepancy naturally becomes still smaller. But here also we find in Sasaram algebraic discrepancy of −8·48 per cent or in Jamui −16·01 per cent and so on. For the total material further compensations between positive and negative discrepancies would occur, but even then a residual discrepancy of + 4·32 per cent had still remained.

254. The absolute percentage discrepancy, on the other hand, is based on the accumulated total positive and negative errors which are not permitted to cancel out. In one sense, this is a better measure of the real magnitude of discrepancies in the two enumerations. For the whole material relating to wheat we notice that it was so high as 74·59 per cent.

255. Similar results of the comparison of the area under gram are given in Table (9/6) in which the arrangement of the columns is exactly the same as that in Table (9/5).

TABLE (9/6). COMPARISON OF CROP RECORDS IN TERMS OF ACREAGE: GRAM

Name of Police Station	J. L. No	Area in acres estimated by		Actual discrepancies (difference between two records)		Sum of discrepancies		Discrepancies as per centage of estimate by first worker	
		first worker (w_1)	second worker (w_2)	positive	negative	algebraic (counting sign of difference)	accumulated (neglecting sign of difference)	algebraic	accumulated
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mohania	134	113·98	120·58	10·95	17·55	−6·60	28·50	−5·99	25·00
	535/1	146·41	130·67	22·22	6·48	15·74	28·70	10·75	19·60
	535/2	140·19	121·54	24·73	6·08	18·65	30·81	13·30	21·98
	537	19·20	18·65	2·32	1·77	0·55	4·09	2·86	21·30
Total		419·78	391·44	60·22	31·88	28·34	92·10	6·75	21·94
Shahpur	146	29·16	24·02	9·14	4·00	5·14	13·14	17·63	45·06
	147/2	49·80	53·20	12·75	16·15	−3·40	28·90	− 6·83	58·03
Total		78·96	77·22	21·89	20·15	1·74	42·04	2·20	53·24
Sasaram	121	113·63	13·02	102·27	1·66	100·61	103·93	88·54	91·46
	122/1	1·93	57·59	0·62	56·28	−55·66	56·90	−2883·94	2948·19
	123	1·06	0·73	0·40	0·07	0·33	0·47	31·13	44·34
	125	0·23	—	0·23	—	0·23	0·23	100·00	100·00
	139	11·71	13·11	3·04	4·44	− 1·40	7·48	11·96	63·88
	140	5·02	3·96	1·89	0·81	1·06	2·72	21·12	54·18
	141/1	45·40	28·32	25·68	8·60	17·08	34·28	37·62	75·51
Total		178·98	116·73	134·13	71·83	62·25	206·01	34·78	115·10
Lakhisarai	125/1	9·97	9·63	6·76	6·42	0·34	13·18	3·41	132·20
Gogri	325/1	1·33	3·09	0·36	2·12	−1·76	2·48	−132·33	186·47
Jamui	10/1	8·99	12·42	3·05	6·48	− 3·43	9·53	−38·15	106·01
	10/2	12·46	14·02	2·96	4·52	− 1·56	7·48	−12·52	60·03
	20	14·91	2·26	14·03	1·38	12·65	15·41	84·84	103·35
	22/2	4·90	9·34	0·33	4·77	−4·44	5·10	−90·61	104·08
Total		41·26	38·04	20·37	17·15	3·22	37·52	7·80	90·94
All Police Stations		730·28	636·15	243·73	149·60	94·13	393·33	12·89	53·86

Here we find that the over-all algebraic percentage discrepancy was 12·89 per cent and the absolute or accumulated discrepancy 53·86 per cent.

256. For convenience of reference the results for each police station are given in a more condensed form in Table (9/7). From col. (8) of this table we notice that the algebraic percentage discrepancy varied from −50·0 per cent in police station Shahpur to −3·1 per cent in police station Lakhiserai with an over-all net discrepancy of +4·3 per cent for wheat. The net discrepancy varied from −132·3 per cent in Gogri to + 2·2 per cent in Shahpur with an over-all value of + 12·9 per cent for gram. The accumulated discrepancies (neglecting their signs) were much higher, and varied from 52·4 per cent in police station Mohania to 180·9 per cent in Shahpur with an over-all value of 74·6 per cent for wheat. The values varied from 21·9 per cent in Mohania to 186·5 per cent in Gogri with an over-all discrepancy of 53·9 per cent for gram.

TABLE (9/7). SUMMARY OF COMPARISONS OF DUPLICATE CROP RECORDS IN TERMS OF ACREAGE : WHEAT

Name of Police Station	Number of		Estimated area under crops in acres		Sum of discrepancies in crop acreage		Percentage discrepancy in crop acreage	
	villages	plots	first survey	second survey	algebraic (counting sign of difference)	accumulated (neglecting sign of difference)	algebraic (counting sign of difference)	accumulated (neglecting sign of difference)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Area under wheat in acres								
Mohania	4	860	97·2	91·9	+ 5·4	50·9	+ 5·5	52·4
Shahpur	2	1608	7·9	11·8	− 3·9	13·9	−50·0	180·9
Sasaram	7	1073	132·7	144·0	−11·2	84·4	− 8·5	63·6
Lakhisarai	1	772	6·4	6·6	− 0·2	9·4	− 3·1	147·1
Gogri	1	50	13·8	11·6	+ 2·1	10·9	+15·4	79·0
Jamui	4	1855	27·9	32·3	− 4·5	43·7	−16·0	157·0
Total	19	6218	285·9	298·2	+12·3	213·2	+ 4·3	74·6
Area under gram in acres								
Mohania	4	860	419·8	391·5	+28·3	92·1	+ 6·8	21·9
Shahpur	2	1608	79·0	77·2	+ 1·7	42·0	+ 2·2	53·2
Sasaram	7	1073	179·0	116·7	+62·3	206·0	+34·8	115·1
Lakhisarai	1	772	90·0	9·6	+ 6·4	13·2	+ 3·4	132·2
Gogri	1	50	1·3	3·1	− 1·8	2·5	−132·3	186·5
Jamui	4	1855	41·2	38·0	+ 3·2	37·5	+ 7·8	90·0
Total	19	6218	730·3	636·1	+94·1	393·3	+12·9	53·9

PART 3. GENERAL REVIEW AND SUMMARY

SECTION 10. GENERAL REVIEW OF METHODS FOR THE IMPROVEMENT OF CROP STATISTICS

257. Agricultural statistics in India especially in the permanently settled provinces like Bengal and Bihar have been known for a long time to be thoroughly unreliable. MacDonnel had complained in 1876 of 'a great want to an administrator in Bengal (which at that time included Bihar), the want of agricultural statistics'. (*Report on the Food-grain Supply of Bengal and Bihar, introduction, p.iii*). Since then nearly half a century has passed and the subject has been reviewed over and over again. The Royal Commission on Agriculture (1928) under the chairmanship of Lord Linlithgow was of opinion that existing agricultural statistics "are admittedly often mere guesses and not infrequently demonstrably absurd guesses" (*Report, p.605*). Dr. A. L. Bowley and Mr. D. H. Robertson were of opinion that "the annual statistics of area are almost worthless" (*Report on a Scheme for the Economic Census of India*, 1934, pp.35–39). The Bengal Paddy Enquiry Committee (1938) fully endorsed these views (*Report, Vol.* 1. *p.* 88). The Bengal Land Revenue Commission (1940) observed that "no dependable statistics existed in Bengal....to show the yield of the various crops" (*Report, Vol.* 1, p.76).

258. Although the need of improvement of crop statistics has been always fully recognized, no definite action, except the five-year sample survey scheme in Bengal initiated in 1937 for the improvement of Jute forecasts and later extended to the paddy crop, had been taken for al ong time. The acute food crisis in Bengal in 1943 and general food shortage in North-East India under war conditions, however, brought things to a head; and large schemes have been and are being sanctioned by Provincial Governments as well as the Government of India for the improvement of crop statistics. It would be useful, therefore, to give a general review of the subject with special reference to permanently settled provinces like Bengal and Bihar in the light of the experience gained in the course of the work on the jute and paddy crops in Bengal from 1937 to 1943 and the pilot survey of *rabi* crops in Bihar in 1943–44.

259. *Subjective and objective methods.* Methods of preparing crop statistics may be broadly classified into two groups, namely, (a) subjective, or (b) objective. Official forecasts in Bengal and Bihar have been so far and are still based on the subjective method. Local officers supply estimates of crop acreage and what is called the 'condition factor' for outturn from a general subjective appreciation of prevailing conditions; and the total produce is calculated by the formula, area × standard yield × condition factor. In the 'Report on the Marketing of Wheat in India' (Government of India, 1937) it is explained (pp.8–9):

> 'In the temporarily settled areas where the basis of assessment for land revenue is revised periodically, e.g., the Punjab, the United Provinces etc., the first factor, namely, area, is known with reasonable accuracy. The second factor, namely, standard yield, is not known with the same degree of accuracy since it is frequently based on old and doubtful data. The effect of seasonal condition is the most difficult to gauge as under the present system reliance is placed on the opinions of the local *patwaries* and village *chaukidar* or *thanadar*.'
>
> 'The estimate of standard yield is understood to represent the average outturn on average soil in a year of average character as deduced from the information obtained from experiments made up to the period under review. Revisions are made from time to time in the standard figures.....'
>
> 'There is no revenue staff in the districts of premanently settled provinces in Bengal and Bihar and the difficulties of estimating the correct seasonal factor are much greater than in the temporarily settled areas.'

After reviewing existing methods the Report states (p.10) 'the general opinion seems to be that little reliance can be placed on estimates based on primary material collected as at present.'

260. Under existing conditions official forecasts based on the subjective or, as it is often called, the reporting, method have been usually under-estimates. This is due to a large extent to the natural desire of the cultivators to make crop estimates lean towards short supply, as under-estimates would help in keeping down assessment of revenue in the temporarily settled areas and also in maintaining higher prices generally everywhere.

261. *Objective methods.* In objective methods estimates of crop acreage would be based on actual physical examination of growing crops and estimates of yield per acre on actual crop-cutting experiments. In this method the condition factor would be completely eliminated, and the annual production would be directly calculated each year by the formula: area × rate of yield per acre. Scientific opinion for a long time has been agreed that this objective (or survey method as it may be called) is to be preferred to the subjective or reporting method.

THE AREA SURVEY

262. *Complete enumeration and sample survey.* So far as the area survey is concerned two different procedures are open, namely, (a) a complete enumeration or (b) a sample survey. In the complete enumeration a complete inventory is prepared by actual physical examination of each individual plot showing the proportion of land under each crop in each plot. In the sample survey similar records are prepared not for all plots but only for a comparatively small number of sample-units located at random over the area under survey. Each method has its advantages and disadvantages. It is not possible to give a categoric verdict in favour of one or the other. In fact, a combination of the two methods, where this can be arranged, may offer the best line of advance.

263. *Comparative merits in the abstract.* The three basic factors which have to be considered in this connexion are (1) the nature of the information, (2) precision of the results, and (3) cost of the operations in two methods. Auxiliary administrative consideration may also be important. It will be useful to start with a brief statement of comparative merits of the two methods in an abstract form.

264. *Nature of information.* The complete enumeration (if it can be carried out without mistakes) would give not only the total area under any particular crop in the province as a whole but would also supply fullest possible information regarding crop acreages separately for districts, sub-divisions, police stations, individual villages, or even individual plots. In wealth of detailed information the complete enumeration thus represents the highest goal of statistics. The sample survey on the other hand can only supply over-all estimates for comparatively large areas like the province as a whole, or individual districts, or even possibly sub-divisions, but not for *thanas* or villages unless the cost is to be disproportionately high. In fact the more detailed the information required the higher will be the cost in the sample survey.

265. *Margin of error.* The total margin of error is made up of three different kinds of components, namely, (a) sampling errors due to fluctuations from sample to sample; (b) recording errors arising from mistakes in the preparation of crop records or at other stages of the work; and (c) margin of uncertainty or variation arising from physical factors. In the complete enumeration the first or (a) sampling errors cannot occur. If we ignore (b) recording or (c) physical errors, then the results of a complete enumeration are theoretically perfect. In the sample survey, on the other hand, the results can never be perfectly accurate. A certain margin of error is inherent in the sampling method because of the fact that the estimate is based not on the entire universe but only on a comparatively small fraction of it picked up at random.

266. *Cost of operations.* Against this, the cost of a sample survey and the labour required are comparatively small. For example , for Bengal , it has been found that a sample survey can give an estimate of the total area under a crop like jute for the province as a whole with a sampling error of something like 3 per cent at a cost of about one tenth or one-fifteenth of that of a complete enumeration. In the abstract , the choice between complete enumeration and sample survey thus depends purely on what information is desired to be collected and what price one is prepared to pay for the same.

EFFECTIVE ACCURACY

267. *Recording errors.* It is however not possible to ignore the recording errors. Mistakes always occur at the stage of recording the anna-estimate of the crop (and also at other stages of the work) in both sample survey and complete enumeration. These arise from chance errors of observation, and also from what is often far more serious, gross negligence or dishonesty on the part of the investigators. Experience in Bengal has shown that such recording mistakes are in practice more important than sampling errors.

268. During the last 7 years we arranged for certain compact areas being surveyed in detail independently by two (or some times three or more) sets of investigators. In this

way a large mass of crop records were prepared in duplicate independently by different sets of workers. A comparison of such duplicate crop records invariably showed large discrepancies between crop entries made by different investigators for the same identical plots. Such material for Bengal has been discussed in various reports submitted to the Indian Central Jute Committee and the Government of Bengal , and also in a memoir on Large Scale Sample Surveys in the *Phylosophical Transactions* of the Royal Society, Vol. 231, No. 584,1944. The proportion of plots for which the records were discrepant was often so high as 20 to 30 per cent.

269. In Bihar, in the *rabi* season of 1944, with investigators receiving Rs. 40/- per month (plus cost of living and travelling allowances on usual scales) and working under close supervision, we found in Section (9) that the proportion of plots for which records in two successive enumerations were discrepant varied from 10 per cent to 15 per cent of the total number of plots surveyed depending upon the stringency of comparison. In terms of acreage we found that, even when positive and negative errors are allowed to cancel out, the over-all residual error was quite appreciable. The absolute or accumulated discrepancy (when the positive and negative errors are not allowed to cancel) was surprisingly high and over 54 per cent for gram and nearly 75 per cent for wheat as shown in Table (9/5). We thus find that the results of a complete enumeration as carried out under existing conditions in either Bengal or Bihar are not entirely free from errors but are subject to appreciable recording mistakes arising from either accidental errors or gross carelessness and shirking of work on the part of the field staff.

270. *Margin of error.* Once it is admitted that the results of a complete enumeration are not free from mistakes, the margin of error of the results requires consideration. In this matter the method of complete enumeration suffers from a great disadvantage. Once any mistake has been made in a complete enumeration there is no possibility (short of a repetition of the enumeration) to detect it. It is therefore not possible (short of a second enumeration) to make any valid estimate of the margin of error of a complete enumeration.

271. *Scientific control in the sample survey.* The sample survey stands on an entirely different footing. Fluctuations in the results must occur in the sample survey owing to the fact that a fraction and not all the units are included in the survey. But these very fluctuations make it possible (by methods based on the theory of probability) to calculate a valid estimate of the margin of error of the final result. This is a distinct advance in scientific method. The sample survey can thus supply not only an estimate of the acreage under a given crop but also the margin of error of the estimate. The prime necessity for this purpose is, of course, that all sample-units which are actually surveyed should be picked up purely at random. Additional controls such as the interpenetrating sub-samples or duplicated enumeration of grids can also be easily provided as explained in previous sections. It is, of course, not possible to guarantee that a sample survey would give accurate results; this must ultimately depend on the reliability of the field records. But controls can be provided in such a way that if a sample goes wrong then this can be detected.

272. *Physical error.* It would be useful at this stage to consider what I have called the margin of uncertainty due to physical causes. The area sown, for example, is not a mathematical quantity which is fixed throughout the season ; crops continue to be sown on new fields till quite late in the season ; plants die for lack of rain or are destroyed by pests or animals. The total area under any crop thus varies from day to day or even from hour to hour throughout the crop season. The area harvested also has its physical margin of error depending on the limit of accuracy with which the area of a field can be determined. For example, there is a good deal of uncertainty in the area of *ails* (field ridges serving as boundaries between fields). Sir John Hubback in his bulletin on "Sampling for Rice Yield in Bihar and Orissa" (pp. 9-10) discussing the harvested area writes. :

> "It........excludes the field ridges (*ails*), which are usually measured in with the fields at a cadastral survey. Since the cadastral survey figure is the basis for the cropped area of the year, this..........point is of importance. No very satisfactory determination of the area occupied by ridges has yet been made, but it would appear to vary beetween nearly 10 per cent in a hilly country to scarcely more than 2 per cent in level country. It is usually taken as 5 per cent."

273. The weight of harvested crop itself is subject to fluctuations due to physical causes. For example, the weight of wheat depends on the humidity of the atmosphere ; according to the "Report on Marketing of Wheat in India" (p. 215) the natural shrinkage may be about "1 per cent between the time the wheat is harvested and the onset of the monsoon"; the gain in weight during the monsoon may be 1% or 2%, or in particularly humid regions even "as much as 5% by the end of September". Whatever be the accuracy of the

results of crop-cutting experiments a good many factors of variation come in at the stage of harvesting, threshing, and winnowing under actual conditions of work in the cultivators' own fields or homes. In considering utilization of the crop other uncertainties are brought in, such as loss during cartage and storage (due to vermins, insects, fungus etc.), wastage in processing and cooking or even as food.

274. In fact, not only crop acreage or crop outturn but every physical quantity has its own margin of uncertainty. For example, the weight of say a grown up man is not something absolutely fixed ; it changes when he takes food or a drink of water, or even when he inhales or exhales air. It is clearly meaningless to try to give the weight of an adult human being correct to say a few grains. In the same way, there must be a certain limit beyond which it is physically meaningless to go in estimating the acreage, the rate of yield per acre, or the total outturn of a crop. In the case of wheat or paddy, the uncertainty due to purely physical factors can be scarcely less than 2 or 3 per cent. To try to reduce the margin of error due to sampling fluctuations below this physical limit is, therefore, meaningless.

THE PATWARI SYSTEM

275. *Temporarily settled provinces.* The lack of accuracy in primary field work discussed above has however reference only to work under existing conditions in Bengal and Bihar. I have no information regarding the accuracy attained in practice in the acreage estimates made by *patwaries* in the temporarily settled provinces like Madras, Punjab, United Provinces etc. It is generally believed that crop estimates prepared by *patwaries* are fairly reliable, and it is quite conceivable that this is so. It must be remembered that the *patwari* is an important link between the village and the Government machinery for collecting land revenue ; he is indeed a man of importance in the village and enjoys great social [illegible] that in practice the office is often hereditary [illegible] knowledge of both cultivators and cultivation conditions. Estimates prepared by the village *patwaries* are therefore much more likely to be reliable than estimates made by *ad hoc* investigators employed from season to season.

276. There is also a second factor of importance. The cultivator is interested in keeping down estimates of production to as low a level as possible for a variety of reasons. In temporarily settled areas there is the question of periodic revision of rent ; a short crop is more likely to keep up prices ; there is the natural tendency to be conservative in estimating chances of gain, of not tempting providence ; and so on. The Government of the province, on the other hand, is interested in collecting as much revenue as possible ; and the *patwari*, as the representative of Government, is interested in increasing his crop estimates to as high a level as possible. It is the balancing of these two forces in opposing directions that make the *patwari* keep to some kind of middle truth.

277. *Permanently settled provinces.* In permanently settled provinces like Bengal and Bihar the *patwari* system does not however exist. To appoint a large number of *ad hoc* investigators, and to give them the name of *patwaries* would not and cannot immediately create a *patwari* system as it exists in the temporarily settled provinces. The *ad hoc* investigator would not have, at least in the beginning, the social prestige of the real *patwari*. He would not have and cannot immediately acquire the same intimate knowledge of village conditions. Finally, the *ad hoc* investigator would not have any connexion with the collection of revenue so that the inter-play of opposing forces noted above would be lacking in the permanently settled provinces.

278. In this situation it cannot be expected that an *ad hoc* staff of investigators would be capable of supplying crop estimates with as much reliability as real *patwaries*. This does not mean that in the course of time a real *patwari* system cannot be built up in either Bengal and Bihar. Four conditions are however essential for success. First of all, workers (whatever they may be called) must be given practically permanent appointments as otherwise Government would not have any hold on them ; and the quality of work would be doubtful. Secondly, they must be given an adequate salary ; it is well known that men on poor pay like village chowkidars cannot be relied upon to collect accurate information. Thirdly, the number of *patwaries* must be sufficiently large to be able to do the work in an adequate manner. Finally, they should be associated with responsible Government work other than the preparation of crop estimates. Linking with the revenue system would of course be the desirable thing ; but, as long as the Permanent Settlement is in existence, this may not be possible. Other suitable work of sufficient importance and prestige should therefore be allotted to make the *patwari* an important link between the village and the Government of the province.

279. Given these four conditions it should be, of course, possible to build up, in the course of time, a really dependable *patwari* system in Bihar and Bengal. From the statistical point of view this would be certainly desirable. This would also have obvious advantages from an administrative point of view. It would be therefore sound policy to organize a *patwari* system in Bengal and Bihar. Two points, however, require to be emphasized, namely, the need of fulfilling the four conditions mentioned above; and secondly, the fact that a real *patwari* system would take an appreciable amount of time to build up. It is, therefore, necessary and desirable in the transitional stage to use the sample survey for obtaining reliable crop statistics. The best line of advance thus lies in the combination of the *patwari* system and the sample survey. Before discussing the advantages of this combined system it will be convenient to describe the essential features of an actual scheme for complete enumeration.

THE BENGAL SCHEME FOR COMPLETE ENUMERATION

280. The Government of Bengal have recently sanctioned a scheme for complete enumeration of acreage under different crops in which work is expected to begin in the *aman* (winter) paddy season of 1944-45. The whole of the province of Bengal is to be covered in the scheme; the total area to be surveyed (exclusive of reserve forests) would be thus about five crores of acres comprising about nine and a half crores of plots in round numbers.

281. *Object of the scheme*. The object of the present scheme is to prepare a complete inventory of all crops grown on each individual plot. The investigators will go round the fields and after actual physical examination of growing crops note against each plot in appro [illegible] the *khasra* lists the anna-proportion (taking the area of the whole plot as 16-anna) of the land. This primary field enumeration [illegible] repeated in each of the three principal agricultural seasons, namely, *bhadoi* (monsoon), *aghani* (wi[illegible]) [illegible] *rabi* (spring). [illegible] entries will also be made of different kinds of fallow land.

282. *Primary investigators*. The province is divided for administrative purposes into 612 Police Stations and 5786 or say 6,000 Unions in round numbers. Since 1941 a large semi-permanent organisation is in existence in Bengal for the issuing of licenses for growing jute under the Bengal Jute Regulation Act. In this Jute Regulation Department (as it is called) there is one primary licensing assistant in each Union, and his part-time services will be utilized for the crop census work. This primary licensing assistant or recorder in each Union will be given the assistance of two literate and active villagers (or *patwaries* as they are being called) for inspecting the plots and noting in the *khasra* or field lists the name and extent of cultivation of each crop against each plot.

283. *Compilation*. The field *khasras* (or crop records) would be sent to compilation offices, one of which would be set up in each circle of six Unions on an average as in the Jute Regulation Department. In these compilation offices the anna proportion of the plot under each crop will be multiplied by the total area of the plot; and the actual area (in acre) under each crop in each plot will be noted. Adding these figures for all plots in a *mauza* the area under each crop would be directly obtained for each *mauza*; and totals for higher units such as *thanas*, sub-divisions, districts etc. would be gradually built up.

284. *Budget*. Each investigator (who is to be called a *patwari*) will be given a consolidated pay of Rs. 100/- per year. The Bengal Scheme states that "this will be a retainer and the *patwari* will live in his own home and be permitted to follow his own pursuits subject to the exigencies of the crop census work. In each Union an area of about 10 or 12 square miles will have to be traversed at least thrice in the year and each plot under crop will have to be inspected". As there are 6,000 Unions, and there will be two village investigators (or *patwaries*) in each Union each on Rs. 100/- per year, the total cost of the primary field staff would be 6,000×2×Rs. 100=Rs. 12,00,000 (twelve lakhs of rupees) per year.

285. For compilation work there will be 1,000 compilers getting between Rs. 50 and Rs. 60 per month; the total cost would be about Rs. 6,30,000 per year inclusive of dearness allowance at 1944 rates. Rent of camp offices would cost Rs. 3,60,000 and contingency and other expenses about Rs. 2,00,000 so that the total cost of the scheme would come to Rs. 23,90,000 per year. This is exclusive of a preliminary non-recurring expenditure of Rs. 3,15,000 in the first year for the preparation of the plot lists. The above estimates of about twenty-four lakhs of rupees per year is also exclusive of the cost of the part-time services of the Jute Regulation staff which will function as the inspecting and supervising staff in the scheme.

286. *Effective time for enumeration*. Two or three points require notice in the above scheme. It is necessary to remember that sowings continue over a considerable span of time

in each season ; and that enumeration of areas in which sowings have not been completed would necessarily lead to under-estimation. The area survey work cannot therefore start in any region until sowings have been practically completed. It is also clear that area enumeration would again become unreliable after harvesting has started. The effective time for area survey work thus lies between the completion of sowings on one hand and the beginning of harvesting on the other. The net time available for area enumeration would not therefore be 12 months but appreciably less, possibly about 7 or 8 months in the year in each region. Even a wholetime worker would thus be able to do area survey work for only from 210 to 240 days in the year.

287. Now in both Bengal and Bihar our experience has been that it takes on an average about 10 investigator-days of work for the complete enumeration of plots in one square mile. On this basis, each Union with an area of about 10 or 12 square miles would require about 100 or 120 investigator-days of work in each season, or from 300 to 360 investigator-days of work for three seasons per year. One investigator per Union is thus not sufficient for complete enumeration work. It is worth noting that in the original form of the Bengal Scheme the proposal was to employ only one local investigator in each Union ; but in the revised scheme, as already noted, two local investigators would be employed in each Union so that, so far as the number of investigators is concerned, the staff would not be inadequate.

288. *Pay of investigators.* The real weakness to my mind lies in their pay and status. Although the annual salary of Rs. 100 is called a 'retainer' in the Bengal scheme, and it is stated that the *patwari* would be "permitted to follow his own pursuits subject to the exigencies of the crop census work", in practice he would have little time left to do other work. The real question then is whether a pay of Rs. 100 per year would attract the kind of person who can be relied upon to do the crop census work with the requisite accuracy. For reasons already explained this appears to be doubtful.

289. *Part-time inspecting arrangements.* Adequate inspection and supervision of the primary field work is indispensable for success. In the Bengal scheme this is intended to be done by the part-time services of the Jute Regulation staff. This may be feasible in the off-season when they are not busy with jutelicensing work ; but it would be difficult to manage this work during the licensing season and at the time of the checking of the licensed area, which means roughly during 6 or 7 months in the year. The plan to get the inspecting work in the complete enumeration scheme done by the part-time help of the Jute Regulation staff may, therefore, turn out to be another weak point in the scheme.

290. *Independent cost.* If a separate inspecting staff has to be provided the cost would come to roughly fifty per cent of the cost of the primary field staff or say about six lakhs of rupees. The total cost of a scheme like the one sanctioned in Bengal with an independent inspecting staff of its own would be thus something like thirty lakhs of rupees per year. If the village investigators (or *patwaries*) have to be paid not Rs.100 per year but a higher salary then the cost would be much greater. For example, if they are given Rs. 25 per month or Rs. 300 per year the additional cost would be 24 lakhs of rupees, or including inspecting cost at a proportionately higher rate say 30 lakhs of rupees. At 1944 level of wages and prices the independent cost of a scheme of complete enumeration like the Bengal one may thus easily come to about sixty lakhs of rupees per year. In any case, if work is done as in the Bengal scheme the cost would be about twentyfour lakhs of rupees per year. To have anything cheaper may mean that the information gathered would be unreliable and not worth collecting.

OUTTURN OF CROP

291. We have been considering so far the problem of the area survey, i.e. of estimating the crop acreage. There is also the problem of estimating the outturn per acre (and hence, by multiplying with the area, the total production) of the crop. Under certain conditions the determination of the yield per acre of a given crop is much easier than estimating the crop acreage ; it is also much less expensive.

HUBBACK'S EXPERIMENTS IN BIHAR : 1923-1925

292. The subject of crop-cutting experiments has received a good deal of attention in recent years. The basic principles were, however, established nearly 20 years ago in India by a series of experiments (which truly deserves to be called classical) carried out in Bihar by Sir John Hubback between 1923 and 1925. He gave an account of his work in a bulletin on 'Sampling for Rice Yield in Bihar and Orrissa' published by the Government of India in 1927. By direct experimental studies on an extensive scale he showed that by an expendi

ture of a sum of money, that was not excessive in view of the importance of the matter, it was possible to replace the normal yield and the condition factor (used in the ordinary method of calculating total outturn) by a scientific estimate in which the possibility of error could be gauged.(p.2).

293. *Need of random samples.* First, as to the need of randomization, he described the existing method and commented on it in the following way (pp.3–4) :-

> 'An officer of some standing visits the tract, of whose yield he is required to form an estimate, and selects a fairly large area, usually 1–10th of an acre, as containing an average crop. This he cuts and carries and in due course threshes and weighs. If he has time he makes two or more experiments, but it is obvious that he cannot usually manage to make a large number of such experiments. There are two main objections to this method : the first is that it depends entirely for its accuracy on the ability of the officer to select. He has to make his selection from a large number of fields growing different varieties at different points of maturity. It is very easy to be misled as to yield by the strength of straw, heavy straw being by no means always correlated with heavy yield of grain. Again it is difficult to give proper weight to the fields which will give little or no yield at all, and even in a normal year such fields are by no means rare. These factors make accurate selection very difficult. The second objection is even more important, viz., that there is no possible way of estimating what is the probability that the result of such selections is within a given range from the true mean yield. The method is comparable to estimating the average income of the population of a town by watching the streets for a few days and then picking out a man, who looked to be in average circumstances and discovering what his income is.'
>
> 'The only way in which a satisfactory estimate can be formed is by as close an approximation to random sampling as the circumstance permits, since that not only gets rid of the personal element of the experimenter, but also makes it possible to say what is the probability that the result of a given number of samples will be within a given range from the true mean. To put this in definite language, it should be possible to find out how many samples will be required to secure that the odds are at least 20 to 1 on the mean of the samples being within one maund of the true mean.'

294. The above statement still remains a masterly discussion of the subject. I have already referred to Yates's paper which fully confirmed Hubback's conclusions. Scientific opinion now is united in insisting upon the need of using random samples. The credit for recognising and emphasizing this principle in India 20 years ago must go to Hubback.

295. *Zoning.* In Hubback's method the work was organized on the basis of subdivisions averaging nearly 1000 square miles in extent. In each such tract 12 centres were to be fixed, spread as evenly as possible over the whole tract (p.8). In each centre the sampler would go out a fixed distance in one direction and circling round return from another direction so as to cover approximately the same area on each day. Crop-cutting work was to be done at each centre for 12 days, the actual dates of harvesting at different centres being spread over the whole period of harvest (p.8). As Hubback points out 'the aim was to distribute the sampling as regularly as possible both in *time* throughout the period of harvest and in *space* over the area sampled' (p.6). At least 20 sample-cuts (which were of quite small size) were expected to be collected per day on an average ; with 12 centres Hubback expected to get 'at least 240 cuttings, but probably as many as 300' in each tract of about 1000 sq. miles. It would be noticed that without making any explicit mention of the fact Hubback had in fact used the principle of zoning in a most effective manner.

296. *Small size of cuts.* As already mentioned in Section 6, Hubback had clearly realized the comparative efficiency of cuts of small size. He had explicitly stated :

> 'It is no advantage to take a large number of samples from places very close together, where the crops will naturally be very much the same on the same day. The degree of accuracy is not seriously improved by such practice. This explains why there is no need to take large samples instead of the handy samples obtained by my method. A sample of one-tenth of an acre is merely 320 of my samples taken in juxtaposition. It simply gives a determination of the mean yield of that particular field, which is not more effectively accurate than that given by say four small samples. Even four samples instead of one are not worth while, because in the great majority of cases they do not differ among themselves enough to affect the mean

> or the standard deviation of the whole set of samples......Technically speaking, there is very high correlation between the individuals of such groups of samples, which makes the ordinary rule, that standard deviation of the mean was the standard deviation of the population divided by the square root of the number of samples, quite inapplicable' (p.9).

This has been fully confirmed by the work of Yates and others already cited in Section 6, and also by our own work on the jute and paddy crops in Bengal. (It is worth noting that Hubback had also recognized the departure of the standard deviation of the mean from 'the inverse square root' rule long before other workers).

297. *Precision of results.* Hubback had given considerable attention to the precision of the results. He had carefully studied the frequency distribution of the sample-yields and found that the standard deviation varied from 4 maunds to 12 maunds per acre (with an average value of about 6 maunds per acre). Hubback pointed out that with 240 cuttings the mean yield of a tract (of about 1000 sq. miles) would be out by only one maund per acre in only one case out of twenty. As each district would consist of about 4 such tracts the above margin of error would be reduced from one maund to about half a maund per acre for individual districts ; and with 16 districts, the margin of error of the mean yield for the province as a whole would be reduced to about one-eighth of a maund or five seers per acre.

298. It is easy to express the same results in the form of percentages. From Hubback's data I find that the coefficient of variation for different *thanas* varied mostly from 30% to 60% with an average value of about 43%. Accepting 43% as a representative figure, we find that the margin of error would be about 5% or 6% for the mean yield of tracts of 1000 sq. miles each ; 2% or 3% for the mean yield of individual districts; and less than one per cent for the mean yield of the province as a whole.

299. It is of interest to recall here a fact already mentioned in Section 6, namely, that the coefficient of variability of mean yield was about 38% or 39% for wheat, and 46% and 67% for gram in the two districts. These compare quite favourably with 43% for paddy found by Hubback. As already mentioned, from 2500 to 3000 cuttings of the handy size used in the present survey (if these can be picked up at random in a suitable manner) would enable the provincial rate of yield with a standard error of about one per cent for wheat, and a little less than two per cent for gram. Hubback's recommendation of using cuts of a small size is thus fully justified by our work on wheat and gram in Bihar in 1943–44.

300. Hubback had emphasized the applicability of his method to small or large tracts. He explained that 'the same number of centres would give a slightly greater degree of accuracy for smaller and slightly less degree for a larger tract, because the standard deviation on the whole increases with the size of the tract, though not at all rapidly' (p.8). Here also he anticipated a good deal of recent findings in this matter. One interesting technical point is also worth noting. Hubback had perceived that his method of randomization did not exactly conform to the requirements of the theory of 'simple' sampling (p.12) or uni-stage random sampling ; he appears in fact to have felt in a vague way the need of using what would be now called the theory of multi-stage sampling.

301. *A special difficulty in crop-cutting work.* There is one peculiar difficulty in crop-cutting work which requires consideration at this stage. The area survey can be done at any time between sowings and harvesting. As this interval lasts for at least several weeks, there is no difficulty in arranging an investigator to reach the village and plots (selected beforehand at random)while the crop is still standing on the field. In crop-cutting work the crop has however to be collected just when it is ready for being harvested ; and this harvesting periotd is usually short and may last for only a few days in each locality. The actual time of harvesting also fluctuates quite widely from one place to another. The result is that, when villages and plots are selected beforehand at random, it often happens that an investigator reaches a particular village when the crop has been already harvested or is yet not ready for harvesting.

302. Hubback had clearly recognized this difficulty. He referred to some work done by Mr. Dobbs (then Director of Agriculture, Bihar and Orissa) in 1921 in Nawada *thana* in Gaya. Dobb's 'method was to divide up the map in equal squares, in order to pick out the villages in which samples were to be taken, and then to sample the fields bearing the survey number 50, or the nearest fields planted with rice'. Hubback pointed out, however, that 'for this method it was necessary to find out the date on which the crop in each field would be ready for harvest and to arrange to visit the village on that date' (p.6) which of course presented great practical difficulties.

303. *Hubback's achievement.* I have dwelt at some length on Hubback's work because of its immediate relevance to the present problem, and also because of what he had achieved nearly 20 years ago at a time when the whole subject was in an undeveloped stage. He had carried out careful experiments extended over three crop seasons (1923–1925) and over the greater part of the province in the Orissa deltaic tract, Chota Nagpur plateau,Santal Parganas and Manbhum, the alluvial Gangetic plain and Gaya; and was of opinion that conclusions drawn from his figures could be 'applied with safety to the whole province except perhaps the North-Gangetic tract' (p.8).

304. He had thus succeeded in working out (with the help of extensive field experiments covering the greater part of a province) a complete method based on the principle of random sampling for estimating the mean yield of paddy with a margin of error of one per cent or less for the province as a whole. He himself had stated that 'the method could also be applied to other crops, especially cereals' (p.9).

TYPE OF FIELD STAFF : STATIONARY OR MOVING

305. *Stationary investigators.* It may be asked, therefore, why could not Hubback's method be directly adopted? This brings us to the crux of the problem, namely the human agency. Hubback had suggested using the 'existing revenue and agricultural staff', and was of opinion that with their help 'there would be no serious difficulty in substituting this (i.e. his) method for the unscientific method of sampling now prescribed' (p.11).

306. The subject may be considered in general terms. Let us assume that the part-time services of a large number, say 1000 persons (to be called 'samplers'), who are scattered more or less uniformly over the whole province, are available for the survey. A number of plots can then be picked up at random in advance and assigned to each 'sampler' in the immediate neighbourhood of his usual place of residence. Each sampler can then keep a watch on the crop growing on the plots allotted to him, and can collect the sample-cuts at the proper time when the crop is ready for being harvested. At full harvesting time there would be no difficulty in each sampler collecting 4 or 5 cuttings on an average in one day. If each sampler works for one day or two at a stretch (or a part of the day for a few days) there would be no difficulty in collecting from eight to ten thousand cuttings which would be quite sufficient for all practical purposes.

307. Three points require to be emphasized in this method. First of all, that each investigator or 'sampler' would do his work in the immediate neighbourhood of his normal place of residence so that he would not have to undertake any long journeys for collecting the samples. This may, therefore, be called the system of '*stationary investigators or samplers*'. Secondly, the crop could be collected just when it is ripe for harvesting so that there would be no wastage of samples or gaps. Thirdly, each sampler would do the work along with his regular occupation, and in fact would put in only one or two days of work so that the cost of field work would be practically negligible, and in any case marginal. We thus find that, given the part-time help of a large number of stationary investigators (like the revenue or agricultural staff of a province), crop-cutting work can be carried out in accordance with the principles of random sampling in a perfectly satisfactory manner at practically negligible cost. Hubback had established this on a firm experimental basis twenty years ago.

308. *Moving investigators.* But, unfortunately, the part time services of a large number of investigators are not always available. For example, in my letter dated 8th November, 1943 addressed to Mr. R. A. E. Williams, I.C.S. (at that time Secretary, Revenue Department, Bihar) I had specifically enquired whether the agricultural and or other Government staff could be made available (of course, on a part-time basis) for crop survey work in Bihar; but I was given to understand that this was not possible and that the work had to be done by investigators appointed under the scheme.

309. When a suitable human agency (like that of the *patwaries* in the temporarily settled provinces) is not, or cannot be made available for the sample survey then there is no other alternative but to get the work done by a comparatively small number of whole-time investigators appointed *ad hoc* for this purpose. As they have to be employed on a full-time basis for the whole season, the area survey has naturally to be spread over the whole season that is, over say three months normally in each crop season. The primary staff required for this purpose is some thing like 250 or 300 investigators. The area commanded by each investigator would then have to be necessarily some thing like 200 or 250 square miles, or, if the double sub-sample plan is adopted, something like 400 or 500 square miles. The investigators cannot any longer be kept stationary but have to move about a great deal to cover

their respective allotted area in an adequate manner. This introduces one complication, namely, the waste of time in moving from one grid or sample-unit to another. The time taken for such journeys (called 'journey time') is by no means negligible, and is usually greater than the total time spent in actual enumeration work. This may be called the system of *moving investigators.* Considerations of cost of field operarions now become of paramount importance; and have been considered in earlier sections.

310. *Possibilities of air-surveys.* I have assumed so far that the primary sample enumeration would be done directly by investigators. It is possible, of course, to use the method of air-survey with the help of photographs taken from aeroplanes. In Bengal this possibility had been considered several years ago in connexion with the jute forecast and actual trial photographs were taken. The difficulty at that time was that the growing jute crop could not be distinguised from other crops. A little later, in 1938, having come to know that similar experiments were being made in U. S. A. I made enquiries but was informed that the difficulties had not been overcome. This subject was again considered by the Jute Census Committee at its third meeting held on the 26th October 1938 (p.5 of preceedings printed by the Indian Central Jute Committee in 1938). My own idea was in fact to try special color sensitizers for this purpose. A grant was actually sanctioned by the Indian Central Jute Committee for photographic experiments; but unfortunately, the beginning of hostilities made it impossible to carry out any work. My personal view has always been that the possibilities of air-survey (using specially sensitized films) deserve serious study, but I regret I have no definite knowledge to give any advice in this matter.

CHOICE OF DIFFERENT METHODS

311. *Choice of the sample survey.* The choice of the sample survey scheme with work done by parties of moving investigators (as in Bihar in 1943–44 or in Bengal since 1937–38) was thus not based on any categoric superiority of the method but because of certain contingent facts. First of all, the sample survey is much less expensive than a complete enumeration. It is capable of supplying, at a small fraction of the cost of complete enumeration, crop acreages for the province as a whole (and possibly districts) with a margin of error of about three per cent. The physical margin of uncertainty, as already pointed out, is also probably of this magnitude. There is, therefore, no point in trying to reduce the sampling error below this limit so that the sample survey is fully adequate so far as provincial (or district) estimates of crop acreages are considered.

312. Secondly, where a *patwari* system (or a similar agency) is not in existence as in Bengal and Bihar it is much easier and quicker to organize a sample survey. A full scheme for complete enumeration in a province like Bengal or Bihar would require an army of several thousand investigators scattered over the whole province. Recruiting and training such a large number of workers is not only difficult but is bound to take a good deal of time. Making adequate arrangements for inspection and supervision (which are essential for success) also present many difficulties. The sample survey, on the other hand, requires only from two to three hunderd primary investigators with an inspecting staff of from 60 to 80 persons for work on a full provincial scale. Problems of organization are therefore much simpler. There is much greater chance of securing say 300 reliable workers than several thousands. Another advantage of the sample survey already noted is the possibility of calculating a valid estimate of the margin of error of the results and of providing controls at the point of collection of the primary material. Finally, in a sample survey it is possible to have progressive estimates of the crop-acreage almost from the begining of the survey and right through the crop season; the final estimate can also be got ready within 10 or 12 days of the termination of the area survey. The administrators can, therefore, maintain practically continuous touch with developments all through. These are considerations in favour of the sample survey.

313. The sample survey, however, has its own limitations. It is unable to supply information for crop acreages for small units of area like *thanas*, villages etc. There are also definite disadvantages in carrying out crop-cutting work by a comparatively small number of moving 'samplers'. As already pointed out, crop-cutting work can be done much better by *patwaries* or some similar agency of 'stationary samplers'.

314. *Advantages of a patwari system.* Where a reliable *patwari* system is in existence it would be in fact possible to utilize its help not merely for crop-cutting work but also for a sample survey of crop acreage. In the area survey, as at present organized in Bengal or Bihar, the total number of investigator-days spent in each season is something like 25,000 in round numbers. As already pointed out, a good deal of this is consumed in moving from

one grid to another and other miscellaneous work. The net time required for actual physical examination of the grids and for preparing crop records is about 12,000 investigator-days. If stationary investigators were available then they would be able to finish the whole field work in these 12,000 investigator-days. With 2000 investigators the whole field work for a full scale sample survey can therefore be finished in about 6 or 7 days.

315. The same thing can be easily seen in a slightly different way. In the *bhadoi* season in Bihar about 60,000 grids (each of size 4-acre) were used for the area survey. With 2000 investigators each would have an allotted area or cell of about 30 or 35 square miles so that each investigator would be able to do the enumeration work within his own cell without leaving his usual place of residence. In these circumstances each investigator would be able to survey 5 or 6 grids per day; 2000 *patwaries* (or stationary investigators of a similar type) would therefore be able to finish 60,000 grids in 6 or 7 days. With 4000 or 6000 *patwaries* the total time required would be proportionately less and only 2 or 3 or 4 days. Even these few days of work could be done either at a stretch or could be spread over a month or more to suit the exigencies of the more regular work of the *patwaries*.

316. A sample survey of crop acreage can be thus carried out very easily with only a few days of work in each crop season by *patwaries* or other workers of a similar type. Crop-cutting work can also be done by them in the same way. Even if only four or five cuts are taken by each sampler per day on an average then work for only a day or two by each of 2000 samplers would enable eight thousand or more sample-cuts being collected for the province as a whole which would be ample for all practical purposes.

317. If *patwaries* (or a large number of stationary investigators) happen to be available then a sample survey of both crop acreage and rate of outturn per acre can therefore be easily carried out with only a few days of work in each crop season. As the primary work would be done by permanent employees of Government its accuracy is likely to be much greater than that of work done by *ad hoc* investigators employed from season to season.

318. The total volume of work in the sample survey is only a fraction of that required for complete enumeration. For a province like Bihar or Bengal with an effective area of say 60,000 square miles to be surveyed (leaving out hilly and forest tracts), at the rate of 10 investigator-days per square mile, the time required for complete enumeration would be 6,00,000 (six lakhs) of investigator-days. For a provincial survey 60,000 grids of 4-acre each would be sufficient, and would require only about 12,000 investigator-days (for stationary investigators); that is, only about one-fifteenth of the time required for complete enumeration. Fairly reliable district figures could be given with say four times as many, or 240,000 grids. Even then the total time required would be only about one-twelfth of that of a complete enumeration. With a system of stationary investigators the cost of a sample survey would be thus extremely small, and in any case marginal and practically negligible. As already noted, the cost of crop-cutting work would be again only a small fraction, one-tenth or one-fifteenth of that of the area survey.

319. *Complete enumeration.* If more detailed information is required about the crop acreage of small units like *thanas* or villages then recourse must be had to the method of complete enumeration. Where *patwaries* are available, a complete inventory of crop acreage can be prepared by them as a part of their routine duties. Where *patwaries* are not avaiable (as in Bengal or Bihar) a suitable system of stationary or local investigators must be organized for this purpose. This, as already pointed out, is partly a matter of expense and partly one of time. Provided enough money is sanctioned for this purpose it would be certainly possible and desirable to build up a *patwarie* system in Bengal or Bihar.

ADVANTAGES OF A COMBINED SYSTEM

320. *Superposed sample survey.* I have already noted that where possible a combined system would offer the best line of advance. In such a combined system, a sample survey of crop acreage as well as crop-cutting work on a restricted scale would be superposed on the work done by the *patwaries* or other local investigators. This would give two independent sets of estimates of crop acreage as also of rate of outturn per acre for the province as a whole or for individual districts; and would supply a most valuable check on the reliability of the results. The superposed sample survey would in fact function in the combined system in a manner analogous to the financial audit in ordinary administration.

321. *Gap between 'area sown' and 'area harvested'.* The sample survey has another important use. The area harvested is not the same as the 'area sown or planted' since practically every year some of the fields bear no crop worth cutting. The harvested area must always be less than the area sown as some of the plants would die or be destroyed by pests

or vermin or animals. In provinces with good irrigation facilities there is naturally much less risk of the crop failing for lack of rain ; and the gap between 'harvested area' and 'area sown' may be quite small. On the other hand, in provinces in which irrigation facilities are meagre there is always a good deal of risk of the crop failing through lack of rain. As it happens, the permanently settled provinces like Bengal and Bihar are also provinces which are backward in irrigation. It is in these provinces that the gap between 'area harvested' and 'area sown' is therefore likely to be quite appreciable. In fact Sir John Hubback after discussing this subject stated that the difference between the planted (or sown) area and the harvested area of paddy might, in a bad year, "well be over 50 per cent all over a sub-division" in Bihar (p. 10 of his bulletin already cited).

322. It is, therefore, clearly necessary to estimate the gap between the 'area sown' and the 'area harvested'. This cannot, however, be done by the method of complete census as enumeration of crop acreage would continue throughout the season. In the sample survey, a quick area survey early in the season would give a good idea of the area sown. A second survey *carried along with crop-cutting work* would supply an objective estimate of the area harvested. This would enable an estimate being made of the gap between the two, and would hence enable suitable adjustments being made in the crop acreage obtained from complete enumeration. The need of estimating the gap between 'area sown' and 'area harvested' is thus by itself a sufficient reason for the continuance of the sample survey along with complete enumeration. A combined acreage and yield survey would also enable the total outturn being estimated directly instead of by multiplying the total acreage by the average rate of yield per acre. This has important theoretical advantages.

323. *Frame-work for sample survey.* In the combined system, when the *patwaries* or local investigators have once become well organized, the sample survey would have chiefly a supplementary function. The scale of operations may, therefore, be reduced to below that at present in use in Bihar and Bengal. But even then in order to organize and carry out the sample survey work it is essential to have a strong frame-work of inspecting and supervising staff. Bengal experience shows that from 60 to 80 inspectors and 20 or 25 chief inspectors would be required for this purpose. If such an inspecting staff is maintained, they would be able to conduct independent area surveys and crop-cutting experiments on a restricted scale which would supply a most valuable check on the reliability of the whole scheme. The existence of such a nuclear staff trained in sample survey work would indeed be a valuable asset to the province. It would be then possible to mobilize this staff at short notice to collect such information relating to rural areas as may be required by Government from time to time.

324. *Safeguard against future set backs.* Maintaining such a nucleus of trained inspecting staff would also be a safeguard against setbacks in statistical matters in future. The food crisis of 1943 showed how Government was hampered at every step for the lack of reliable crop statistics. This, coupled with conditions brought about by the war situation in North East India, made Government willing to spend large sums of money for collection of crop statistics. In the face of impending famine and pestilence, expenditure howevter large, is quite rightly, of no consideration. But if by any mischance complete enumeraion does not work quite satisfactorily or has to be abandoned at any time in future the one possible alternative would be, of course, to fall back on the sample survey. This, however, would be only possible if a nucleur inspecting staff is available ready at hand for this purpose.

325. The cost of maintaining a skeleton frame-work of sample survey will be small. On the scale visualized above, the inspecting staff may be maintained as an integral part of the complete enumeration (or *patwari* system) without involving any additional expenditure. The real additional cost of the sample survey would thus be the expenditure required for the statistical portion of the work. On the reduced scale indicated above, the total cost would probably be something of the order of a lakh or a lakh and a half of rupees per year. This would be a very moderate expenditure for ensuring continuity of work in future. It would also supply independent estimates based on the sampling method and furnish an excellent check on the accurary of complete enumeration.

SECTION 11. SUMMARY OF THE REPORT

326. For convenience of reference a general summary of the Report is given below.

SECTION 2. METHOD OF THE SAMPLE SURVEY

327. A general account is given in this section of the method of the sample survey as developed in the course of a five-year scheme (1937-1941) for the improvement of the acreage estimates of the jute crop in Bengal, and later extended to the paddy crop by the Government of Bengal in 1942, 1943 and 1944.

328. *Use of random samples.* The essential feature is to use about sixty or seventy thousand sample units called 'grids' which are small units of square shape and usually of size from 2 to 5-acre each. The whole of the province is divided into a suitable number of zone-cells. (In Bengal, 960 zone-cells of 64 square miles each are being used at present ; in the *rabi* survey in Bihar, 96 zone-cells each of about 50 square miles were used in each of the two districts Shahabad and Monghyr) A certain number (depending upon the design of the survey) of such grids are located purely at random on cadastral survey *mauza* maps within each zone-cell. Field lists (*khasra*) are prepared showing the revenue serial number of all plots falling within each grid. Investigators then go out to the field with these lists and appropriate C. S. village maps, identify the plots, examine the crops growing on the plots, and record in the field lists the anna-estimates of the proportion of land under different crops on each plot. The field lists are then sent to the Statistical Laboratory where the records are tabulated, and estimates are prepared of the crop acreage based on the information collected for the sample plots.

329. *Planning of the sample survey.* The precision of the results naturally depends on the size (area) of each individual grid and their total number. The larger the size and or the larger the total number of grids the greater is the precision. The object of efficient planning of a large scale sample survey is to settle the size of each individual grid, and their total number and distribution among different zones in such a way as to obtain the final result with the least possible margin of error.

330. *Margin of error.* The margin of error of the results of a sample survey is made up of two parts ; one due to fluctuations of sampling (which is inherent in the sampling method and which can be calculated on the basis of the theory of probability), and the other due to mistakes in the primary records arising from careless or even dishonest work on the part of investigators. Under existing conditions recording mistakes are quite as important as sampling fluctuations. It is, therefore, necessary to incorporate in the sample survey suitable statistical checks and controls.

331. *Sub-samples.* One type of control has proved extremely useful in Bengal. The total number of grids in each zone is divided into two equal portions, say (A) and (B). Grids belonging to each set are located at random all over the area in each zone-cell, and information is collected for the two sets of grids by two different parties of field investigators who work independently and at different times in the same zone so that they never meet. The two sub-samples thus supply two independent estimates of crop-acreage for each zone and also for higher units like districts or the province as a whole. How far these two estimates based on the two sub-samples are in agreement with each other immediately supply a good idea about the reliability (or otherwise) of the results.

332. *Crop-cutting experiments.* The yield per acre has to be determined by direct crop-cutting work in which the crop grown on a suitable measured area is harvested and directly weighed. Extensive experimental studies in India and abroad have shown conclusively that the attempt to select so-called representative fields always introduces bias. Scientific opinion in every country is united on the need of the sample-cuts being located in a purely random manner without leaving any choice to the investigators.

SECTION 3. THE DESIGN OF THE SURVEY

333. *Aim.* The object of the pilot survey in the *rabi* season was two-fold, namely, to obtain some information for the *rabi* crops in the two important wheat districts of Shahabad and Monghyr ; and secondly, to collect basic information necessary for efficient planning of the sample survey in subsequent seasons.

334. *Size and density of grids.* Partly for lack of adequate time of preparatory work and partly to avoid complications in the exploratory stage it was decided to use grids of a square shape and one uniform size, namely four-acre. It was also decided to have, on an average, a density of two grids per square mile.

335. *Zone-cells and sub-samples.* Each of the districts was divided into 96 zone-cells roughly of a square shape and from 40 to 50 square miles in area ; and from 80 to 100 grids were allotted to each zone-cell. For lack of time a detailed inter-penetration of grids was not attempted. The 96 zone-cells in each district were however divided into 48 pairs of adjoining cells. Among these 24 pairs of zone-cells were allotted at random to sub-sample (A), and the remaining 24 pairs to sub-sample (B). One party of field investigators worked in each zone-cell so that the information was collected independently for sub-samples (A) and (B) for each pair. With detailed inter-penetration of grids a comparison between the two sub-samples would have been possible for each zone-cell separately. In the present design, valid comparison can, however, be made only for the district as a whole.

336. *Sub-blocks.* Each district was divided into three blocks and each block into two sub-blocks each consisting of 16 zone-cells. Each block was placed in charge of a chief inspector under whom there were four field parties. Each field party consisted of four investigators under one inspector. Of the two parties in each sub-block one party belonged to sub-sample (A) and the other to sub-sample (B). *

337. *Complete enumeration.* Arrangements were also made for the complete enumeration of all plots in certain compact areas ; portions of these areas were subsequently surveyed independently by a second party. The object of collecting such plot to plot information was two-fold : firstly, to carry out special studies by model sampling experiments, and secondly, to furnish some idea as to the degree of reliability of the primary field work.

SECTION 4. GENERAL ACCOUNT OF THE SURVEY

338. *Preparatory work.* Planning of the survey started in the third week of December 1943, and was completed within a few days. The work of locating at random 18,000 sample units or grids on appropriate village maps, preparing field lists (*khasra*) comprising more than two lakhs of individual plots and making other necessary arrangements for the survey was completed by the end of the first week of February 1944.

339. *Organization of the field branch.* Seventy-four experienced crop survey workers in charge of two field supervisors were sent to Bihar on loan from Bengal. The local field staff in Bihar, recruiting of which started in January 1944, consisted of 74 investigators, 22 inspectors and 6 chief inspectors. A training camp was opened at Gogri in Monghyr district on 3rd February 1944, where the newly recruited Bihar staff was given actual training in field work by the inspecting and supervising staff sent on loan from Bengal.

340. *Progress of field work.* Field work started on 12th February and terminated on 13th April 1944. The total period of work was 62 days out of which 2 days were treated as holidays on account of *holi* festival ; the actual period of 60 days of work was divided into 5 haltages of 12 days each. A special conference was arranged at Bikramgunj and was attended by the whole of the supervising staff and technicians. Various problems relating to mixed crops and crop-cutting experiments were discussed and decisions relating to future procedure were made.

341. *Tabulation and statistical analysis.* Arrangements were made for despatching the material collected in each haltage directly to Giridih. The original plan was to have the whole field data tabulated within a week or so of the receipt of the material. This plan however had to be changed owing firstly to the delay in receiving the field records during the first two haltages, and secondly because of the need of starting preparatory work for the early *makai* (maize) survey which was decided to be taken up in May. A further change in plan was necessitated by the decision to organize a full scale survey in the *bhadoi* season for which preparatory work had to be started on a very large scale.

342. *Reports.* A preliminary report on the survey in the *rabi* season based on material collected in 3 (out of 5) haltages was submitted on the 28th April 1944. Partly due to heavy pressure of work in connexion with the *bhadoi* survey and partly to the sudden illness of the Statistical Adviser the writing of the present report was unduly delayed, and was completed at the end of October 1944.

CHIEF RESULTS OF THE RABI SURVEY

343. The pilot survey in the *rabi* season, although it was confined to only two districts, has supplied a great deal of valuable information relating to crop statistics.

SECTION 5. THE AREA SURVEY

344. The present survey has shown that a sample survey of this kind can supply district estimates of acreage under wheat with a sampling error of about five per cent and a

provincial estimate with a sampling error of one or two per cent. For gram the sampling error for the province as a whole would be something like two or three per cent (Section 5).

345. *Estimated crop acreages.* In Shahabad district the sample estimates of crop acreage (in thousand acres) are 300·35±19·87 for wheat, 78·40±9·51 for barley, 337·33±26·86 for gram, 56·60±12·95 for *arhar* and 197·81±23·78 for *khesari*. In Monghyr district crop acreages (in thousand acres) are 284·08±31·62 for wheat, 66·90±5·32 for barley, 193·43±26·86 for gram, 103·71±23·22 for *arhar* and 47·80±15·11 for *khesari*. The estimated margin of error generally varied between 10 and 30 thousand acres for the different crops in the two districts. This is reasonably small and is quite satisfactory for a pilot survey.

346. In the area survey the internal agreement between different sub-samples is on the whole satisfactory showing that the results are quite reliable.

347. Owing to the late start of the survey, acreage figures are, however, probably under-estimates as portions of the crop had been already harvested during the later stages of the area enumeration. Adjustments based on the observed decrease in successive haltages indicates higher acreages (in thousand acres) of 335, 102, 452, 73 and 298 for wheat, barley, gram, *arhar* and *khesari* in Shahabad district ; and 296, 76, 255, 132 and 70 respectively for the five crops in Monghyr district. These adjusted values are probably upper limits for the area sown.

348. *Settlement figures.* Settlement figures are generally believed to be reliable and may be accepted as representing actual conditions at the time of settlement operations (1907-1916 in Shahabad and 1909-1912 in Monghyr). Crop acreages according to Settlement Reports were (in thousand acres) 228, 145 for wheat ; 105 and 81 for barley ; and 340 and 211 for gram in the two districts respectively. Sample estimates are not strictly comparable to official figures owing to the ambiguity in the interpretation of acreage under mixed crops. But a general examination of the sample estimates against the background of Settlement figures suggests that in recent years the cultivation of wheat has increased appreciably, the area under barley has decreased somewhat, while that under gram has remained much the same in both Shahabad and Monghyr districts. The area under *arhar* is certainly much greater than 1000 acres in Monghyr district as given in official forecasts.

SECTION 6. OUTTURN OF WHEAT AND GRAM

349. *Hubback's work* Sir John Hubback had carried out an extensive series of crop-cutting experiments on paddy covering the greater part of Bihar and Orissa about 20 years ago (1923-25) using, as far as is known, for the first time in India, the method of random samples for this purpose. He had used cuts of a handy size of 1/3200 acre (=13·6 sq. feet) each, and had shown that such small cuts gave satisfactory results. It appears that by his method with 3000 sample-cuts the mean yield of paddy for the province as a whole could be determined with a sampling error of about one per cent (Sections 6 and 10).

350. *Yield per acre of wheat and gram.* Hubback had expressed the opinion that his method would be easily applicable to cereals. This has been confirmed by the crop-cutting work on wheat and gram done in the present survey. The adjusted average yield of wheat grain was 6.48 and 6.45 maunds per acre for Shahabad and Monghyr respectively. The rate of yield of gram (extracted from pods) was 4.98 and 4.25 maunds per acre for Shahabad and Monghyr respectively. The percentage standard error is estimated at 4·7 per cent and 4·8 per cent for wheat in Shahabad and Monghyr respectively and 5·9 per cent for gram in Shahabad and 9·1 percent in Monghyr district.

351. *Margin of error.* The work done shows that about 3000 sample-cuts of a convenient size (30 square feet per plot) located at random in an appropriate manner would enable the provincial mean rate of yield per acre being determined with a sampling error of about one per cent for wheat and a little less than two per cent for gram. It is clear, therefore, that given a suitable human agency for doing the crop-cutting work there is no difficulty in determining the provincial or even district rates of yield with sufficient precision for all practical purposes (Section 6).

352. *District outturn.* The total outturn, in theory, can be obtained simply by multiplying the crop acreage figure by the mean rate of yield per acre. The estimated outturn of wheat (in thousands of maunds of extracted grain) calculated in this way is 1947·1 in Shahabad and 1832·1 in Monghyr ; and of gram 1681·6 and 1822·4 thousand manuds in the two districts respectively.

353. *Unreliability of official estimates.* Comparison of above estimates with official figures is not possible for two reasons. As already noted, acreage figures for mixed crops

are ambiguous. Secondly, official figures relating to rates of yield, 'condition factor', and outturn are conflicting and of doubtful value. Actual calculations based on official reports for the four crop-seasons from 1940-41 to 1943-44 show that "standard" rates of yield have changed suddenly from year to year although no crop-cutting experiments on *rabi* crops appears to have been done by the Department of Agriculture in recent years. Different sets of condition factors are given for the same crop in the same year, but nothing has been stated to reconcile the discrepancies. Enquiries made from departmental officers made it clear that the present position is unsatisfactory, and careful experimental study is essential for progress.

SECTION 7. PROBLEM OF MIXED CROP

354. *Ambiguity about mixed crops.* Mixed crops may have two different meanings. In the same plot one portion of the land may be sown with one crop, and another portion with another crop and so on ; in such cases there is no ambiguity in the area under each individual crop. It is also possible for the same portion of land to be sown with a mixture of two or more crops. In this case each crop covers the whole geographical area under mixed crops. If each crop is credited with the gross area then the total land recorded to be under cultivation would be much greater than the actual geographical area. Estimates may also be made of the net proportion of land which have been covered by each crop if it had been sown separately ; the total net area would be now equal to the geographical area covered by mixed crops.

355. *Frequent occurrence of mixed crops.* Sample studies show that wheat or gram is sown by itself (or 'pure' as it is called) in only a small proportion (about 10 per cent) of the land under wheat or gram alone ; in the remaining 90 per cent wheat (or gram) is sown in mixture with other crops.

356. *Variations in mixture.* In Bihar the relative proportions of the different crops in mixture do not remain the same but fluctuate widely from region to region or even from plot to plot in the same region. Proper assessment of area under mixed *rabi* crops is thus a matter of great importance in this province. As far as could be ascertained, no standard procedure is being used at present for this purpose in Bihar. The Settlement Reports also fail to make the position clear.

357. *Variations in rate of yield of mixed crops.* The rate of yield of wheat was determined for plots with different compositions of wheat sown in mixture with other crops. The rate of yield increases, as one would expect, with increasing composition of wheat in the mixture. But the increase in yield does not take place in equal proportion with the increase in the intensity of wheat. For example, the graduated yield is 4.32 maunds per acre for one and two-anna proportion of wheat, and is 7·4 maunds per acre for a 15 and 16--anna proportion. Thinner compositions thus give relatively higher yields which probably explains their popularity.

358. *Gross acreage and calculated outturn.* In the present survey estimates had been made on the basis of net or effective acreages. Converting these net or effective acreages into gross acreages and multiplying by rates of yield appropriate to different compositions the estimated total outturn in thousand maunds of wheat (as extracted grain) was 3860 for Shahabad and 4486 for Monghyr against calculated values of about 1947 and 1832 as given in Section 6. The wide gap shows the urgent need of critical studies.

359. It is suggested that the best plan would be to record two things in the case of mixed crops, namely, (a) the gross (or geographical) area under all combined crops in the mixture, and (b) the effective or net anna proportion of composition of each crop, that is, the estimated proportion of the land which would be covered by the crop if sown by itself without mixture. Crop-cutting experiments should also be conducted to determine the rate of yield per acre (on the basis of gross or actual geographical area) for plots with different proportions of the crop in the mixture. The outturn can be then calculated directly by multiplying the gross area by the rates of yield appropriate to the particular anna-composition of the crop.

SECTION 8. COST OF OPERATIONS

360. *Volume of labour used in the survey.* In the present survey payment was made to investigators for a total of 6564 days of work out of which 4565 were spent in area survey (inclusive of 317 in crop-cutting work), and 1999 in special studies. The amount of time required for crop-cutting work thus represented only about 7 per acent of the time required for area survey work. The sampling errors of estimates of crop acreage and of rates of yield

per acre were about the same; this shows that area survey work requires about fourteen times more than crop-cutting work to attain the same precision in the results.

361. *Relation between cost and density of grids.* It is found that, as one would expect, the time taken for enumeration of grids increased when the density of grids was small that is, when there were fewer grids per square mile because proportionately more time was consumed in moving from one grid to another. The actual relation between cost and density of grids was similar to that previously observed in Bengal.

362. The above studies were directly useful in preparing the plan for the survey in the *bhadoi* and succeeding sesaons as it enabled calculations being made in advance of the time (and hence expenditure) which would be required in working with grids of different densities.

363. *Analysis of costs.* The total expenditure in field branch as well as the proportionate cost per investigator-month under different heads in the *rabi* survey have been analysed and compared with corresponding costs in the Bengal crop survey in 1943–44. Expenditure and cost per investigator-month in the statistical branch have been also discussed.

SECTION 9. ACCURACY OF AREA ENUMERATION

364. *Recording mistakes.* In certain compact areas complete inventories were made of the proportion of land under each crop in each individual plot. This work was repeated independently a second time in certain areas by entirely different parties of workers. A comparison of such duplicate records for the same identical plots showed that the two sets of record were discrepant in the case of from 10 per cent to 15 per cent of plots compared depending upon the stringency of the definition of agreement.

365. Comparison in terms of acreage showed that both positive and negative discrepancies (i.e. errors of both commission and omission) occurred in appreciable numbers. In building up totals for higher units like villages, *thanas*, districts etc. the positive and negative discrepancies would cancel out to a large extent, and the residual discrepancy would become gradually smaller. For the material under comparison this residual discrepancy was, however, still appreciable and greater than 4 per cent. If the discrepancies are added together neglecting their signs one gets the accumulated discrepancy which was surprisingly high,—about 75 per cent for wheat and 54 per cent for gram.

SPECIAL STUDIES

366. A number of special studies were also undertaken to collect information for improving the technique of the crop survey; and some of the results are mentioned below.

367. *Weight of crop as harvested.* Studies were made to find the relation between the ears of wheat as harvested and the weight of grain extracted from the dried ears as also the relation between the weight of pods of gram as harvested and the weight of extracted grain. If a stable ratio can be found between these two weights then it would be possible to simplify the crop-cutting work by recording only the weight of ears of wheat (or pods of gram) as harvested, and later converting these into equivalent weight of grain. Preliminary results were encouraging as the observed ratio was found to be fairly stable but further investigations are necessary. Preliminary studies were also made of the effect of premature harvesting on the weight of crop (section 6).

368. *Analysis of variance.* There was considerable variation in the yield of both wheat and gram from one region to another and also from plot to plot. Valuable information relating to such variations has been collected which indicates that sample-cuts should be scattered over as many localities as possible rather than on many plots within the same locality, and within each locality on as many plots as possible rather than on many cuts within the same plot. With the help of numerical data calculated from experimental observations it is possible (in conjunction with considerations of costs) to arrange the pattern of sampling in such a way that the highest precision in results may be attained with a minimum expenditure of time or money (Section 6).

SECTION 10. REVIEW OF METHODS FOR IMPROVEMENT OF CROP STATISTICS

369. Section 10 gives a general review of methods for improvement of crop statistics with special reference to permanently settled provinces like Bihar and Bengal.

370. *Existing methods unreliable.* Existing methods of preparing crop statistics are based on subjective estimates made by local officers of both crop acreage and the so-called 'condition factor' of outturn. In the permanently settled provinces figures collected in this way have been known from a long time to be unreliable.

371. *Need of an objective method.* Scientific opinion is united regarding the need of using an objective method based on actual physical examination of growing crops.

372. For acreage surveys two procedures are open, namely, (a) complete enumeration and (b) sample survey. In the complete enumeration a detailed inventory is prepared by direct physical examination of the proportion of land under different crops in each individual plot in the province.

373. For crop-cutting work, complete census of weighing of the crop on each individual plot is out of question, and recourse must be had to the method of random samples.

374. *Sample survey.* In this method a suitable number of sample-units or grids are located purely at random, and the proportion of land under different crops in the plots included within these grids is recorded by direct physical examination. Estimates are then prepared for the area under survey on the basis of such sample information.

375. *Stationary investigators.* The procedure of the sample survey itself takes two different forms depending upon the kind of human agency which may be available for the field work. If the part-time services of a large number of investigators (like *patwaries* in temporarily settled provinces) happen to be available then each investigator can be allotted a small compact area or cell so that he can enumerate a suitable number of grids within his own cell without leaving his usual place of residence. This may be called a system of stationary investigators. If the part-time help of say, 2000 investigators happen to be available then the whole work of area survey can be finished in 5 or 6 days in each season ; with a large number of investigators the work would take still less time.

376. Crop-cutting work becomes extremely simple with a system of stationary observers. A suitable number of villages and plots can be selected at random beforehand and assigned to each investigator within his own cell so that he can collect the samples just when the crop on the selected plots happen to be ready for harvesting. With one or two thousand investigators the whole work can be finished in a day or two practically at negligible cost.

377. *Moving investigators.* If the part-time help of stationary investigators is not available (and it may be mentioned here, that this was not available in the Bihar scheme) then the work must be done by a comparatively small number of whole-time investigators appointed *ad hoc* for this purpose. The area allotted to each investigator in this system must necessarily be large, several hundred square miles in extent. Each investigator has ,therefore, to move about from place to place for both area survey and crop-cutting work. A good deal of time is consumed in such journeys, and considerations of cost become important in deciding the pattern or design of the sampling technique.

378. *Advantage of complete enumeration.* A complete enumeration (when it can be carried out without mistakes) represents in wealth of detailed information the highest goal of statistical science. Against this, the sample survey can only give crop acreages for large units like the province or districts or perhaps sub-divisoins, but not for small units like *thanas* or villages.

379. *Margin of error.* Once, however, any mistake is made in the complete enumeration there is no possibility of detecting it except by a second (or possibly a third) enumeration. The sample survey possesses a great advantage in this matter ; provided the sample-units are taken at random, it is possible to estimate the margin of error on a valid and scientific basis. It is also possible in the sample survey to incorporate internal statistical checks and controls. Although a sample survey cannot be guaranteed to give accurate results, if it goes wrong, this can be detected.

380. *Relative costs.* The cost of a complete enumeration is necessarily high. Experimental studies in both Bihar and Bengal show that about 10 investigator-days are required per square mile to do the work satisfactorily. On this basis, 6,00,000 (six lakhs) of investigator-days would be required for the complete enumeration of 60,000 square miles. For provincial estimate of crop acreage the use of about 60,000 grids of about 4-acre each would be sufficient for which the time required for actual enumeration would be about 12,000 or 15,000 investigator-days at the most. With stationary investigators a sample survey of area would thus require one-fortieth or one-fiftieth of the labour of complete enumeration. With moving investigators more time would be required but even then it would be something like one-twentieth of that of complete enumeration.

381. *Recording mistakes.* In actual practice, one difficulty in the method of complete enumeration under existing conditions in Bengal and Bihar is that large mistakes are made in the preparation of crop records and also at other stages. For crop acreage, the effective

accuracy of a complete enumeration therefore is not likely, in the immediate future, to be higher than that attained in the sample survey in practice. Organizational difficulties in working with a large staff are also naturally greater in the complete enumeration.

382. *Choice between the two systems.* The choice between complete enumeration and sample survey is thus primarily a question of expense, and depends entirely upon in how much detail information is desired to be collected and how much expenditure and trouble one is prepared to incur for this purpose.

383. *The patwari system.* I a real *patwari* system can be built up (or where it is in existence) it should be possible to get quite reliable estimates of crop acreage. This system has distinct advantages over a sample survey (carried out by a small rapidly moving staff) in crop-cutting work. A *patwari* system has also many administrative advantages; but in order to be effective it must have a sound organization as in the temporarily settled provinces.

384. *A combined system.* In Bengal and Bihar the building up of such a *patwari* system will take time. In the transitional stage it is desirable to continue the sample survey in order to secure reliable crop statistics in the immediate future.

385. Even when a *patwari* system is fully organized it would be desirable to incorporate the sample survey as an integral part of the scheme of complete enumeration. In such a combined system both sample survey of area and crop-cutting work would be done with the part-time help of the *patwaries* for which only a few days of work would be required in each crop season.

386. *Gap between area sown and area harvested.* The area harvested is always less than the area sown. The gap may be large in a province like Bihar in which crops are usually rain-fed; according to Hubback this gap may be so large as 50% for paddy in large tracts of Bihar. In a combined system, an early sample survey of acreage would be carried out in the middle of the season to estimate the 'area sown'; and a second sample survey of area would be carried out *along with crop-cutting work* to estimate the 'area harvested'. It would be then possible to make necessary adjustments for the gap between the two. This combined acreage and yield survey would also enable the total outturn being estimated directly for the grids and zones and hence for districts or the province as a whole.

387. *Sample checks.* The sample survey would in any case supply valuable checks on the results obtained through complete enumeration.

388. *Marginal cost.* In the combined system it would not be necessary to have a separate body of investigators for the sample survey; but a small trained staff of inspectors would be maintained for general supervision of the work. The cost of a sample survey integrated with a scheme of complete enumeration would be thus marginal and small, probably of the order of a lakh and a half of rupees per year in a province like Bihar.

389. *Safeguard against future set-back.* The maintenance of a nuclear staff for sample survey would be useful in case by any mischance the scheme of complete enumeration is not entirely satisfactory or has to be abandoned in future as it would be then possible to fall back on the sample survey for securing crop statistics of sufficient accuracy for practical purposes.

ADDENDUM: EDITORIAL NOTE

In recent years the Indian Statistical Institute had undertaken a large number of sample surveys of various kinds on behalf of Govenment Departments in India and had submitted report on such surveys from time to time. Publication of these reports under war conditions was not considered advisable. Hostilities having now ceased it is intended to publish such of the reports as are likely to be of interest to applied statisticians. It is not proposed to follow a chronological or any other definite sequence. The general aim would be to give concrete examples of the type of work which is being done by the Institute and to indicate the wide scope and possibilities of the method of the sample survey, and the reports would be printed, as far as possible, in the form in which they were submitted.

17th August, 1945

P. C. Mahalanobis,

INDIAN STATISTICAL INSTITUTE

Twelfth Annual Report: 1943-44.

We are writing this report in the midst of renewed enemy action in North-East India. The year under review has been one of unprecedented stress and strain in Bengal. The country passed through an acute food-crisis of a kind never known before. In consequence there has been repidly growing demand of statistics in every direction. Although we were obliged to expand our staff at a rapid rate, the growth was not altogether healthy as there was a great scarcity of trained statistical workers. The only thing we could do was to train up as many men as we could in the course of the wark itself.

During the year under review the total number of worker-months in employment on the statistical side was 2472 or an average of 206 workers per month. The statistical work itself had to be carried out at three places, in Calcutta, at Baranagore in the suburbs of Calcutta, and at Giridih a little over 200 miles away from Calcutta. The statistical staff was divided roughly in equal proportions in the three centres at the beginning, but, owing to lack of accommodation in the Presidency College in Calcutta, the number of workers at both Baranagore and Giridih had to be steadily increased so that early in 1944 it was something like 150 or 160 at each of these two places. It is only because of the unsparing loyalty and enthusiasm of the old guards of the Statistical Laboratory that it has been at all possible to carry on the work at such high pressure.

Crop Survey work in Bengal

The most notable progress in the year under review was achieved in the sample survey of crops. It would be remembered that work was started on a small scale with a total expenditure of about Rs. 7,000 on an exploratory survey of the jute crop in Bengal in 1937. This was the beginning of a five-year scheme for the improvement of jute statistics which was financed jointly by the Indian Central Jute Committee and the Government of Bengal. The sampling technique. developed in the course of a gradually expanding series of surveys culminating in the full scale provinicial work of 1941, which demonstrated beyond dispute the possibility of securing by this method a final estimate of jute acreage with a margin of error of only two or three per cent at a cost of one-fifteenth or one-twentieth of that of a complete enumeration. A general account of the five-year scheme with discussion of theoretical foundations was given in a memoir prepared by the Hon. Secretary in 1942 which is being published in the *Philosophical Transactions* of the Royal Society of London.

The Hon. Secretary had been pressing from the very beginning of the five-year scheme in 1937 for the extension of the method to cover paddy and other important crops in Bengal. Each year from 1938 to 1942 the repeated his efforts but without success. In 1942 the Indian Central Jute Committee had expressed their complete satisfaction with the method of the sample survey and had recommended its continuance by Government in future. The Government of Bengal however decided at first to discontinue the work but subsequently revoked their decision and asked the Hon. Secretary at the end of March 1942 to proceed with the Jute survey. By this time a good portion of Burma had been occupied by the Japanese, and the supply of rice from that country had been cut off. Apprehending a serious deterioration in the food situation (not because of the physical volume of the import which was small but because of the possible effect on prices) the Hon. Secretary submitted a definite scheme to the Government of India at the end of March 1942 for extending the sample survey to cover the paddy crop in Bengal. In the course of the next few months he also discussed the question with a large number of officials in Bengal but all his efforts failed completely. In the absence of reliable statistics both the provincial and the Cental Governments were left entirely in the dark regarding the supply position of rice at the end of 1942. This made it impossible for Government to pursue a consistent food policy on any objective basis. The weakness of the statistical position was thus an important factor in bringing about the deplorable food crisis in Bengal in 1943. The Hon. Secretary has given a general account of the position in a paper presented before a symposium organized by the National Institute of Sciences of India held in Calcutta at the end of September 1943 which has been already printed in their Proceedings Vol. X, No. 1 (1944).

Early in 1943 the Hon. Secretary again submitted a scheme to the Government of Bengal for a sample survey of all important crops in Bengal. Sanction was given to that portion of the scheme which related to the jute crop and preparatory work was started at the end of February 1943. On the basis of informal discussions the Hon. Secretary also proceeded with the preparatory work for the survey of paddy which however had to be discontinued in the middle of June 1943 under instructions from the Secretary of the Bengal Agricultural Department as it was not considered advisable by Government to proceed with the scheme. In August 1943 the Hon'ble Mr. H. S. . Suhrawardy, Minister in charge of the Department of Civil Supplies, decided however to take up the survey from his own Department and issued orders to proceed with it. In this way we got the opportunity for the first time in 1943 to undertake a comprehensive survey of the paddy crop in Bengal covering about 70,000 square miles.

Design of the Survey : Work in 1942 had to be carried out under great difficulties in consequence of which the quality of the field work had deteriorated somewhat. Conditions in the country in 1943 continued to be, if anything, evern more difficult on account of the acute food crisis. Many of the field staff were raw recruits and had no previous experience of this kind of work. We, therefore, decided to strengthen the statistical control by securing duplicate enumeration of about one-sixth of the grids. In order to do this the uni-stage method had to be given up. Adjoining police stations were paired, and out of each pair one was selected at random thus including in the sample 50 per cent of the police stations from all over the province. Within each selected police station a certain number of sheets of maps were selected at random ; and within each sheet a suitable number of grids were located again at random. In view of the fact that jute as well as paddy had to be covered in the survey it was decided to use one uniform size of grid, namely, 2.25 acre for the whole province ; and the total number of such grids actually ennumerated was about 59,000.

Field Organization : Eight training camps were opened in February 1943 and about 225 workers of different ranks were given field training. The total field staff numbering 365 was organized early in April and the area survey of jute and Aus paddy was started in the East Bengal districts by the middle of May 1943 and was extended to other districts in the course of a few weeks. The area survey was completed by the end of August but crop cutting work on jute and aus paddy continued till the middle of September. The area survey of Aman paddy began early in October and was finished by December 1943 crop cutting work started by the middle of November and continued till January 1944.

Precision of Estimates : The area under Aman paddy a based on the sample survey was found to be in large excess of the official figure of 179 lakhs of acres. The field records were therefore carefully scrutinized. Although there was some evidence of the field work not having been done quite honesty in certain areas, as far as we could judge, these defects were not sufficient to explain the large difference. In this situation we made a careful comparison of the acreage under Aman paddy as given in Settlement Reports with the contemporaneous official forecasts and found that during the period of 41 years from 1900 to 1941 the official estimates were on an average 24 percent below the corresponding Settlement figures. In the light of this comparison the sample estimate was quite plausible.

Crop-cutting work in the Jute-Aus season could not be made fully representative of the different regions. In the Aman season wark was done at 120 centres in 25 districts. The margin of error (at odds of 1 to 20) of the estimated total production o rice (not in husk) in the case of aman paddy was of the order of 7 or 8 per cent. Such a large margin of error is of curse unsatisfactory, but must be attributed mainly to the unreliability of the primary meterial especially in crop cutting work. Here difficulties of inspection and checking are very great and everything depends on the honesty of the investigators. Unfortunately, in 1943 we were obliged to work with 303 new recruits (or 64% per cent of total field staff of 476 men) who had no previous experience of the work and about whom we did not know anything. It is not surprising, therefore, that certain portions of the material were found to be unreliable. No remedy is possible unless and until the Government of Bengal adopt the policy of building up a more permanent staff of investigators.

BENGAL CROP SURVEY : 1944-45

A scheme for a sample survey of Jute, Aus and Aman paddy in Bengal in 1944-45 was approved by the Government of Bengal early in 1944 and a consolidated block grant of Rs. 4,20,000 has been sanctioned for this purpose. In order to improve the efficiency certain modifications have been introduced in the design of the survey. The whole province has been divided into zone-cells of practically equal size of 64 square miles (8miles-square parallel to the latitude and longitude lines) and a two-stage sampling techinque is being used on the basis of sub-cells of size one square mile each. The preparatory work is progressing smoothly at the time of writing this report. The Government of Bengal have placed at the disposal of the Institute the part-time services of Mr. N. C. Chakravarti, B.C.S., (Addl. Asst. Secretary, Revenue Department) as Superintendent of Statistical Surveys : and the Field Branch is being actively supervised by him.

CROP SURVEY IN BIHAR

At the request of the Government of Bihar the Hon. Secretary went to Patna on the 16th November 1943 and discussed the possibility of organizing a sample survey of crops in the province with H. E. Mr. R. F. Mudie, Governor of Bihar, and other officials. A five-year scheme at an estimated cost of twentyfive lakhs of rupees was submitted on the 19th November and the main proposals were accepted in principle by the Government within a few days. This visualized a series of gradually expanding surveys beginning from the *rabi* season of 1943-44 and reaching a full scale provincial survey in the *aghani* (winter) season of 1945.46. In this scheme the field work would be done under the direct control of the Government of Bihar for which a new post of whole-time Superintendent of Statistical Surveys was created in December 1943 and was filled by a senior and experienced Deputy Collector, Rai Bahadur S. P. Sinha, in December 1943. The Indian Statistical Institute would remain responsible for the statistical work during the first five years, but would gradually transfer the work to the Government of Bihar.

The preparatory work for the *rabi* season was started in Calcutta and Giridih in December 1943. The first design, which took a good deal of time and labour to prepare had to be abandoned owing to the delay in receiving village maps from different districts. At the end of December 1943 the programme was revised ; it was decided to take up the exploratory work in the two districts of Monghyr and Shahabad ; and a new design was got ready for this purpose and the preparatory work was pushed on with all possible speed at Giridih.

Arrangements were made to send a staff of 72 field workers from Bengal on loan to Bihar and 108 men were recruited locally for the work. The field survey started on the 12th February and was closed on the 18th April 1944. Two special camps were opened at Bikramgunj in Shahabad district and Gogri in Monghyr district, and a number of trained statisticians were posted to these two places for special studies. The Hony Secretary inspected the work at Bikramgunj on the 25th and 26th Mzrch 1944 and a conference of statistical and field workers was arranged there during his visit.

The field survey in the *rabi* season was organized with such success that it was decided to do the work on the full provincial scale beginning from the next *bhadoi* (autumn) season. This means a stepping up of the programme by about 16 months and would naturally involve a very heavy pressure of work in the near future. Preparatory work has been already started for the *bahdoi* (autumn) season of 1944-45, and it is hoped that the field survey for the early crop of *makai* (maize) would be started in the first week of May, and of the other regular crops of the season in July.

A block grant of about Rs. 62,000 was received from the Government of Bihar up to 31st March, 1944, and a budget of Rs. 2,48,000 has been submitted for the statistical portion of the work in 1944-45. The Bihar Government also sanctioned a contribution of Rs. 15,000 per year with effect from 15th November, 1943, in lieu of the services as Statistical Adviser of Prof. P. C. Mahalanobis who had expressed a desire that such a contribution should be made instead of paying him any remuneration.

BURDWAN-HOOGHLY-HOWRAH IRRIGATION SCHEME

This is the sixth year of the scheme of crop-cutting work in this area. In 1938-39 the individual plots of land in this region were classified in terms of certain standard soil classes, and the object of the crop cutting scheme is to assess and grade the different classes of soil in terms of fertility as measured by the rate of yield of paddy. During the first four years the whole of the crop cutting work was done under the direct control of Government. In 1942 the work was done for the first time under our technical guidance and in 1943 crop cutting material was collected along with similar work done in the Bengal Crop Survey Scheme. A comparison of the material collected in 1942 with those collected in earlier years has revealed glaring discrepancies. It is now proposed to make a detailed scrutiny of the primary data with a view to ascertaining whether the material collected in earlier years in sufficiently reliable to be included in the analysis or whether it will have to be rejected altogether.

OTHER PROJECTS AND SPECIAL ENQUIRIES

Besides the crop-area and yield estimation surveys described above, work was also done on other projects and special enquiries. The more important among these are mentioned below.

Regional survey of Borer Pests of Sugarcase 1940-41 : The remaining portion of reduction of the primary material and preparation of the final tables included in the scheme of analysis approved by the Imperial Council of Agricultural Research, was completed during the current financial year. Most of the computational work for this purpose was conducted at Giridih. The report for Gurdaspur region (Punjab) was submitted to the I.C.A.R. by the middle of November 1942. The remaining regional reports, namely those for Muzaffarnagar (United Provinces), Gorakhpur (United Provinces), Dacca (Bengal), Cuttack (Orissa), Gudiyattam (Madras), Padegaon (Bombay), Sepaya (Bihar) and Pusa (Bihar) were completed in quick succession within the next two months. Each report included more than 200 pages of tables dealing with the incidence of damage to Sugarcane from different kinds of Borer Pests.

U. P. Anthropometric Measurements : All calculations leading to the mean values, standard deviations and coefficients of variation with the standard errors of these constants were completed for thirteen characters and five indices for each of 23 castes from the United Provinces and were sent in November 1943, to Dr. D. N. Majumdar of the Lucknow University who had taken the measurements at the time of the 1941 Census. All the correlation coefficients have been determined and advanced work is now being carried on at Calcutta including the use of the D^2 -statistic.

River Research : A grant of Rs. 5000 per year has been sanctioned by the River Research Institute of the Government of Bengal for statistical analysis of data relating to rriver physics, rainfall and floods in Bengal. Work was started in November 1943 on the analysis of gauge readings of the river Damodar severe flood in which during the monsoon season had completely interrupted communications in the East Indian Railway near Burdwan for more than two months. A preliminary report on maximum floods near Burdwan was submitted in December, 1943.

Blood Pressure Survey 1935-36 : Analysis of the large body of material collected during a survey started in 1935-36 in Bengal was taken up this year and the work is now nearing completion.

Sriniketan Rural Survey : Work on the analysis of a large volume of material relating to rural families collected by the Visvabharati Institute of Rural Reconstruction at Sriniketan was taken up at Giridih.

THEORETICAL RESEARCHES

Theoretical work was continued on the following main topics : (1) Design of Large Scale Sample Surveys ; (2) Design and Analysis of Experiments ; (3) Multivariate Distributions ; and (4) Construction of Numerical Tables. A list of papers published and reports submitted in the year under review by the workers associated with the Institute is given in Appendix 1 below.

Large Scale Sample Surveys : This topic is receiving increasing attention from statisticians and Government Departments, as it provides an efficient yet cheap alternative method to that of surveying by complete enumeration. A paper embodying the findings of the theoretical and experimental arches on this topic written by P. C. Mahalanobis will be shortly published in the *Philosophical Transactions* of the Royal Society. A paper on multi-stage sampling was communicated to the 31st Session of the Indian Science Congress (Delhi) 1944.

Design and Analysis of Experiments : Further progress has been achieved in the mathematical theory of the design and analysis of experiments by R. C. Bose, K. R. Nair, C. R. Rao and H. K. Nandi. A general method of approach for recovering the intra and inter-block informations from the results of a general incomplete block design and the necessary tests of significance have been discovered. Confounded balanced designs in the case of symmetrical factorial designs s^m in block of s^k in plots $(k>m)$ have been constructed. Analysis of co-variance has been extended to the case of a general incomplete block design. New types of designs in quasi-Latin squares with efficiencies higher than that of incomplete block designs have been discovered. Considerable progress has been made in finding out the number of non-isomorphic solutions in the case of incomplete block designs.

Multivariate Distributions : In the year under review S. N. Roy has obtained suitable reduction formulae by which it is possible to evaluate numerically the incomplete probability integrals (both on the null and the non-null hypotheses) of the individual p-statistics. These statistics had been introduced by him for purposes of discrimination in respect of variances and co-variances between two multi-variate normal populations, or in respect of mean values of k multivariate normal populations supposed to have the same set of variances and co-variances : the individual distributions of these p-statistics had already been worked out by him in the preceding year. It may be further mentioned that Mahalanobis's Generalized Distance (D^2) is finding suitable application in discriminating between and classifying different castes and tribes in the United Provinces on the basis of anthropometric measurements made during the 1941 Census.

Construction of Numerical Tables : Very useful work on the construction of mathematrical tables, and the unbiased partitioning of the tail areas of various important statistical distributions including those for χ^2 and t, (under non-null hypotheses) is being continued by P. K. Bose and S. Raja Rao under the general guidance of P. C. Mahalanobis. Calculation of 5% and 1% points of the classical D^2 -statistic for different values of the number of variates is nearing completion. A table of Bessel Functions for different values of an imaginary argument has been prepared and will be soon published.

STATISTICAL ENQUIRIES

A large number of enquiries were attended to through correspondence this year, as in previous years. The distribution by subjects and administrative regions of the more important enquiries are shown below :—

By Subjects : Agricultural 9 ; Anthropological 2 ; Educational 7 ; Industrial 19 ; Sample Survey 4 ; Socio-economic 3 ; Theoretical 7 ; Vital 1 ; Miscellaneous 1 ; Total 53.

By Regions : Assam 1 ; Bengal 15 ; Bihar 1 ; Bomaby 16 ; C. P. and Berar 2 ; Delhi 5 ; Madras 1 ; Orisa 4 ; Sind 2 ; U. P. 6 ; Total 53.

ARRANGEMENTS IN THE STATISTICAL LABORATORY

As mentioned in earlier reports, portions of the Statistical Laboratory were removed as an evacuation measure in 1942 to Giridih and Baranagore ; and the work of the Institute had therefore to be carried on at three places, namely, Calcutta, Baranagore and Giridih. The Hon. Secretary went to Giridih several times during the year, and other workers were also sent there from time to time. Owing to increasing postal delays and difficulties of communication arrangements had to be made for important records being carried between Giridih and Calcutta by special messengers. The number of workers in Calcutta remained more or less stationary on account of lack of accommodation, but the staff at Baranagore and Giridih had to be steadily expanded to cope with the increasing pressure of work.

NUMBER OF WORKERS

The number of workers employed on a monthly basis was 81 and the number of piece-rate workers 17 on the 1st April 1943. During the year under review 331 new appointments were made on a monthly basis and 142 as piece-rate workers while 141 on monthly basis and 130 piece-rate workers left the Laboratory so that on the 31st March 1944 the year closed with 271 workers on a monthly basis and 29 piece-rate workers with a total strength of 300 workers in all. The average strength of workers was 80 in Calcutta, 61 at Baranagore, and 65 at Giridih during the year under review.

INSTITUTE STAFF

K. R. Nair, Birendra Nath Ghose, Jitendramohan Sen Gupta and Mrs. Chameli Bose worked in Calcutta most of the time ; N. T. Mathew was in charge of the Giridih Branch during the whole of the year under review. Rajchandra Bose, Samarendranath Roy and Purnendu Kumar Bose although working as whole-time teachers in the Post-graduate Department of Statistics of the Calcutta University, maintained their intimate contact with the Statistical Laboratory and shared fully in its activities. In the Computing Section, Nimaicharan Ghose and S. Raja Rao assisted by Sambhu Nath Haldar, Arun Kumar Ganguli, Suresh Chandra Sarkhel, and Shyamsundar Bose, were in general charge of the work in Calcutta and Baranagore, while Jitendra Nath Taluqdar with Haribhajan Chaudhuri assisted by Dhirendra Chandra Sarkar looked after the computing work at Giridih. Among old workers Sudhir Kumar Banerjee, Heremba Chandra Ghosal, Subodh Chandra Das Gupta, Jaladhar Sarma, Gayanath Adak, Sarat Chandra Mitra, Probhat Chandra Sinha, Debabrata Dutta, Sukumar Hazra, Purna Chandra Ghosh, looked after various section in Calcutta, Baranagore and Giridih. M. K. Dutt has been appointed as Assistant Secretary of the Institute from 1st February 1944, in place of K. Gupta who has resigned. In the Field Branch, Dhirendra Mohan Ganguly was in charge as Supervisor and Pranay Kumar Chatterjee and Santipriya Bose acted as Assistant Supervisors.

Women's Section in Calcutta : A women's section was started in Calcutta after the successful experiment at Giridih. The women workers, about fifteen in number, are being paid on the piece-rate basis. An organizer has been appointed and put in charge of this section. Some of these workers take their work home and do it there, and some work regularly under the supervision of the organizer.

Changes in staff : During the year under review a good number of our workers (both senior and junior) left the Institute to take up more remunerative appointments elsewhere. K. R. Nair, who has been with us for about seven years, and K. Gupta, Assistant Secretary, joined the Food Department of the Government of India, early in 1944 both as Assistant Directors of Statistics. Anil Bhattacharya, another senior worker, joined the Department of Commerce of the Bihar Government in December 1943 as statistician. Many of the technical apprentices also left us. Sushil K. Banerjee joined the Office of the Controller of Rationing of the Bengal Government in Calcutta as statistical officer in November 1943 ; Manindra Nath Ghose went to the Dacca University in November 1943 as lecturer in statistics ; V. Ramachandra Rao joined the Swastika Oil Mills, Bombay in January 1944 as statitician ; Promoda Nath Chowdhury joined the Labour Commissioner's Office at Patna in December 1943 ; Bibhutosh Sircar joined the Calico Mills at Ahmedabad in January 1944 ; M. B. Rudra joined as Deputy Superintendent of Central Excise ; Abani K. Mukherjee (a student of the post-graduate department of Statistics, who had just passed his M.A. examination in statistics in the first class) joined the Office of the Sugar Controller of the Government of India at Simla in September 1943. Mr. Tu, a post graduate Research Fellow from China, who had joined the Laboratory in June 1942 left us in February 1944 to take up some war work in his own country.

Against these departures, several young persons have joined our Laboratory either as technical apprentices or in other capacities ; C. Radhakrishna Rao, Harikinkar Nandi and Nirmalkumar Chakravarti stood 1st, 2nd and 3rd respectively in the first class in the M.A. and M.Sc. examinations in Statistics of the Calcutta University held for the first time in 1943. Miss Bina Sen is a first class first M.A. in Mathematics of the Dacca University and had Statistics as an optional subject in the degree examination ; Mrs. Sri Rangamma Muthanna is an M.A. in Mathematics with Statistics as a special paper of the Mysore University ; Ramkrishna Mukherjee is a first class first M.A. in Anthropology of the Calcutta University. A list of the technical apprentices is shown in Appendix 3.

Branch Laboratory at Giridih

The year 1943-44 was one of steady expansion at Giridih. The number of workers started with 43 in April 1943 and increased to 155 in January 1944. N. T. Mathew, Statistician, was in charge of the Giridih Laboratory throughout the year, and carried on his work with great energy, enthusiasm and ability.

In April 1943 the staff consisted of men who had been transferred from Calcutta as an evacuation measure in 1942 together with a few workers recruited locally. This was gradually strengthened chiefly by local recruitment on a piece-rate basis about which further details are given below. In October 1943 the Hon. Secretary together with a large number of statisticians and other workers arrived from Calcutta and stayed at Giridih for more than one month for advanced work and research of various kinds. In January 1944 the staff had to be expanded very rapidly for the preparatory work for the Bihar Crop Survey Scheme the bulk of which was done at Giridih ; a general description has already been given of this work in an earlier section. The peculiar feature of the arrangements at Giridih is the fact that a large number of the workers live in premises in the occupation of the Institute. This has made it possible to make various experiments in organization and has led to an appreciable increase in co-operative activities of many kinds.

Piece rate System : At Giridih systematic efforts were made to work out a flexible scheme for payment by piece-rate system. This was introduced in the beginning partly to meet a fluctuating demand for computing work and partly to ensure equitable remuneration being given to more painstaking and careful workers. In actual practice a standard rate of output has been prepared (and is revised from time to time) for a large number of individual items of computational work, and definite rates of payment are fixed for each item. Payment however is not made only on the basis of output, accuracy is also taken into consideration and suitable deductions are made for mistakes in calculation. The system has worked successfully and it has been found possible to provide work on piece-rate basis for from 50 to 75 per cent of the total number of workers.

Women's Section : One advantage of the piece rate system is the possibility of supplying certain items of work to women who could not keep regular hours, but could devote a certain amount of time during leisure hours at home. In the beginning all the women workers used to work at home, but gradually a number of them began to attend the Laboratory regularly at first on a piece-rate basis and later on some of them attended as whole-time workers.

Training Section : At Giridih great difficulties had to be faced due to shortage of trained hands especially in January when the staff had to be more than doubled all on a sudden from 72 to 155 in order to cope with the Bihar Crop Survey work. This led to the initiation of organized training courses for computers, and such training was given to about 30 workers including three women in February and March.

Library : As mentioned in earlier reports the main library and valuable records were removed to Giridih early in 1942 as an evacuation measure and have been stored there for the present. From Giridih nearly 4,000 pages of photoprints of books and articles were sent to Calcutta for the use of workers there. The preparation of a classified catalogue with author and titles was taken in hand and was progressing smoothly. In spite of limited resources a circulating-library of general literature was started at Giridih in 1942 and has been working successfully.

Bihar Crop Survey : The bulk of the preparatory work of the Bihar Crop Survey was done at Giridih. This was started early in December. The work had to be expanded rapidly in January and about 50 workers were trasferred from Calcutta to Giridih for this purpose and a large number were recruited locally. At the end of January the total number of workers at Giridih was about 155 among whom more than 90 had to be provided with both food and lodging. Lack of accommodation was a great difficulty. At the instance of the Government of Bihar, the Deputy Commissioner, Hazaribagh, placed at out disposal buildings and premises belonging to one Mr. Myers situated very close to the premises at present occupied by the Institute. This however is not adequate, and an application has been already submitted for the acquisition of land near the Institute premises and as soon as possession is obtained of this land it is proposed to put up additional buildings.

Other activities : The Hon. Secretary during his stay at Giridih in October 1943 gave a series of 12 lectures followed by discussions on the history of statistics with special reference to the work of Karl Pearson and his school. A large number of statistical workers and university teachers who were present at Giridih attended these meetings which often continued for two or three hours at a time. At the end of October 1943 the workers of the Statistical Laboratory at Giridih staged Rabindranath Tagore's play *Vaikunther Khata* in Bengali in aid of the Bengal Famine Relief Fund : the Hon, Secretary took the part of Vaikuntha. A sum of over Rs. 1,200 was collected and has been handed over to famine relief organizations.

BRANCH LABORATORY AT BARANAGORE

Early in 1942 it had been decided to remove a portion of the Statistical Laboratory to Amrapali, the Hon. Secretary's garden house at Baranagore, as a safety measure against the contingency of our being requested to vacate the rooms in the Presiency College, Calcutta. In 1943 owing to expansion in our activities the computing staff had to be steadliy increased, and, as no further accommodation was availabe in the Presidency College, more and more men had to be transferred to Baranagore from Calcutta so that during the preparatory work for the survey of paddy in Bengal in 1943 the total number of workers increased to over 100 at Baranagore. Even this was not enough, and a new shed for computing work was built in October 1943 which increased the accommodation appreciably. At present the bulk of the primary computational work of the Bengal Crop Survey is being done at Baranagore ; and over 3,00,000 C.S. village maps are stored there for this purpose. The Central Store of the Institute has been removed there, and a separate office has also been organized. The total number of workers at Baranagore increased to over 150 in February 1944. A number of workers live in the premises with the Hon. Secretary and Mrs. Mahalanobis ; they all have food from a common kitchen run by the workers themselves. Two other messes for workers are run in renteed houses in the neighbourhood and provide accommodation for about 40 workers. Under war conditions the Institute has been thus obliged to run residential establishments for workers at two places, Giridih and Baranagore.

Summer Vacation Camp at Kalimpong : During the summer vacation of 1943 a Camp Laboratory was set up at Kalimpong for about three months and was attended among others by Prof. P. C. Mahalanobis, Rajchandra Bose, Samarendranath Roy, Birendranath Ghosh, Purnendukumar Bose, Jitendramohan Sen Gupta, S. Raja Rao, Monimohan Mukherjee, Arun Kumar Ganguli, Probhat Kumar Sinha, Kanailal Indra : about 7 or 8 computers were also recruited locally. Sudhir Kumar Banerjee and the three Field Supervisors in the Bengal Crop Survey Scheme, Dhirendramohan Ganguli, Santipriya Bose and Pranayakumar Chatterjee also went there for a short time. A good deal of progress was made in research work on both theoretical and applied topics, the workers being comparatively free from the routine administration and outside engagements of the Calcutta Laboratory. It may be mentioned here that such vacation camps were being organized almost continuously from 1929, at first informally up to 1937, and since then in a more formal manner from the Institute beginning with the summer camp at Darjeeling in 1938.

STATISTICAL LIBRARY

The bulk of the stock of the library continues to be maintained at Giridih in accordance with the general policy adopted for the duration of the war. For the convenience of the workers a branch is maintained in Calcutta, and a branch was opened at Baranagore in the year under review. The library functioned normally and books were requisitioned from Giridih for Calcutta and Baranagore as necessary. This was supplemented by the use of the micro-film photo record apparatus by which photographs were made of articles and books on cinema films and sent to Calcutta.

The total number of books added to the library during the year was 440. Of these 394 books were purchased; 31 were received as free gifts ; and 15 books were received for review in Sankhya. Even after four years of war we received 59 annuals and 161 periodicals. The number of volumes issued by the lending section was 1347 and the number issued by the reference section was 2568, the total being 3915 (as against 2231 in 1942-43.)

The Librarian, Jibananda Saha, who was sent for training to the Imperial Library (Government of India) passed the Diploma examination with credit. A reorganization scheme was adopted for making the Library arrangements up-to-date and scientific. An Assistant Librarian, Biswajit Roy, trained at the Imperial Library and Bengal Library Association (in co-operation with Calcutta University) was appointed to speed up this scheme. Systematic work has been taken up in classifying the books in accordance with the Dewey Decimal classification system and preparing (a) a Dictionary and (b) a Classified catalogue. A current bibliography of Statistics based on the stock in the library is also being prepared. Besides the main scientific library, a circulating library of general literature is also being maintained for the benefit of the workers.

SANKHYA : THE INDIAN JOURNAL OF STATISTICS

Due to the war emergency it was not possible to bring out last year more than two issues (Vol. 6 Parts 3 and 4). In some cases there was a most deplorable lag between the date of receipt of a paper and its publication. As a result of a series of discussions among senior workers a strong group of workers under the leadership of Mr. Raj Chandra Bose have undertaken to bring up to date the technical editing of all papers in hand at the earliest opportunity.

The audited accounts are given below :—

Sankhya : Receipts and Payments Account for the year ending 31st March, 1944

RECEIPTS	Rs.	A.	P.	Rs.	A.	P.
To Opening Balance						
Cash in hand ..	41	4	3			
Cash at Bank ..	294	1	9	335	6	0
,, Subscriptions				340	0	0
,, Grant from the Government of India				2,000	0	0
,, Advance from Statistical Publishing Society.				300	0	0
				2,975	6	0

PAYMENTS	Rs.	A.	P.	Rs.	A.	P.
By Allowance ..				150	0	0
,, Press Bills and Paper				2,408	11	9
,, Contingent Expenses ..				23	5	3
,, Closing Balance :						
Cash in hand ..	2	5	6			
Cash at Bank ..	390	15	6	393	5	0
				2,975	6	0

TRAINING WORK AND EXAMINATIONS IN STATISTICS

During the war emergency it is natural that officers would not come for training or consultation, specially from outside Bengal, so freely as in the pre-war days. Still a fair number of them came here for training or scientific consultation and a complete list of such officers is shown in Appendix 2. K. R. Nair was in charge of this section up to 6th February 1944, and Birendra Nath Ghosh from that date. As regards the standard training course run by the Institute, which also suffered owing to the war emergency, we had six students in all, but mostly they stayed for a shorter period than the full session of one year as they were either absorbed in the Institute staff or they accepted employment elsewhere (Appendix 3).

Institute Examinations in Statistics : In spite of the war emergency, Computer's Certificate Examinations were held from 13th to 17th August 1943, in which 19 candidates appeared of whom 9 were successful. The Board of Examiners consisted of Professors K. B. Madhava, U. S. Nair, P. C. Mahalanobis and Messrs. R. C. Bose, S. N. Roy and K. R. Nair. It is proposed to hold these examinations again in the near future.

VISITORS

As in previous years a large number of visitors came to the Statistical Laboratory during the year under review. Among them was Prof. A. V. Hill, F.R.S., M.P., Secretary, Royal Society of London, who looked into the work in the Calcutta Laboratory and also spent several hours at Baranagore where he saw the arrangements made for the large staff engaged on the Bengal Crop Survey Scheme. Sir C. D. Deshmukh, Governor, Reserve Bank of India and one of the Vice-Presidents of the Institute saw the work both in Calcutta and at Baranagore at close quarters. Another distinguished visitor was Dr. Shing Shen Chern a member of the Chinese Academy who had come to India for a short visit. During the year under review contacts were also established with a number of Army Officers interested in the application of Statistics to various subjects.

LOCAL BRANCHES OF THE INSTITUTE

Up to the time of going to the press, we received the reports of the following branches, and short summaries of their activities are given below.

Delhi Branch : The Delhi Branch, though only in its second year of existence, has been carrying on its activities with energy. Thanks are due to the authorities of the Hindu College, who are allowing the Branch to hold its meetings in the College premises. Election of office bearers and members of the Executive Committee took place on 26th October, 1943 with Sir Shri Ram in the chair. Sir T. E. Gregory, Economic Adviser to the Government of India, was elected President, Dr. P. V. Sukhatme Treasurer, and Mr. G. R. Seth, Secretary of the Branch. The following three lectures were arranged during the year, all of which were well-attended and were followed by lively discussions :—

(1) Statistical Engineering in Business Organization by Mr. M. V. Divatia.
(2) A Study on the Comparison of Milk by Dr. P. V. Sukhatme.
(3) Large Scale Sample Surveys by Mr. K. R. Nair.

Lucknow (U. P.) Branch : During the year 1943-44, this branch has been conducting an investigation into the standard of living and social conditions of the middle classes in Lucknow City. Two investigators have been appointed and have collected data from 500 families by random sample method in respect of the size of the family, diet, fertility, income, literacy, housing, sickness etc. The data collected are now being analysed and a report will be ready shortly. Prof. Radhakamal Mukerjee was in charge as Secretary of the Branch.

Poona (Deccan) Branch : Prof. V. G. Kale continued to be the Chairman of the Branch. Prof. D. G. Karve who ably filled the office of the Local Secretary for the last 12 years, laid down his office in January 1944, and Mr. N. V. Sovani was elected as the new Secretary. The number of life-members increased to 12 this year. The members of the Branch were frequently consulted by official and non-official agencies in connexion with rationing, price control, procurement of food grains, etc.

Mysore Branch : The Branch worked, as usual, in collaboration with the departments of Statistics, Economics, and Psychology of the University of Mysore. Five meetings were held in the year at which papers were read and discussed. An exhibition was arranged at which computing machines, charts and diagrams, schedules of certain statistical enquiries and portraits of leading statisticians were on view. Prof. K. B. Madhava was Secretary during the year under review.

MISCELLANEOUS

Post-Graduate Department of Statistics : Our collaboration with the post-graduate department of Statistics of the Calcutta University continued as before and the Hon. Secretary was the Hon. Head of the Department throughout the year. Purnendu Kumar Bose officiated as whole-time Assistant Lecturer for about three months in place of Sadasiv Sen Gupta (a former senior worker of the Institute) who had taken leave from the University in August 1943 to join work as Statistician to the East Indian Railway ; Purnendu Bose is now acting as part-time lecturer in place of K. R. Nair who left for Delhi in 1944. Mukunda Chandra Chakravarti from Dacca (who had been working in the Statistical Laboratory during vacations since 1942) joined as Asst. Lecturer in December, 1943. Birendra Nath Ghosh joined the department in November, 1943 and is acting as part-time lecturer in place of Anil Bhattacharya who left for Patna in December 1943. C. Radha Krinshna Rao and Harikinkar Nandi (who stood first and second respectively in the first class in the M.A., and M.Sc. examinations in August 1943) are acting as part-time Assistant Lecturers. Mrs. Chameli Bose and Nirmal Kumar Chakravarti (who also had passed the M.Sc. examination in the first class) are working as Honorary Lecturers.

The examination of the Calcutta University for the M.A. and M.Sc. degrees in Statistics was held for the first time in August 1943; and five students passed it after completing a two-year whole time course. It may also be mentioned that the students of the post graduate statistics department together with the teaching staff actively participated for about a fortnight in November 1943 in the field survey especially the crop-cutting work on Aman paddy at two of the camps of the Bengal Crop Survey at Baniban (Howrah) and Singur (Hooghly).

Loan Services : We place on record our grateful appreciation of the courtesy shown by the Indian Central Jute Committee by lending us the services of Mr. Dhirendra Mohan Ganguli, and by the authorities of the Visva-Bharati by lending us the services of Mr. Santipriya Bose for the greater part of the year for supervision work in the field branch of the Bengal Crop Survey Scheme.

Photo record apparatus : We offer our thanks to the Royal Asiatic Society of Bengal for having kindly agreed to keep in our custody their photo-record camera with which it would be possible to take photographs of papers and documents on cinema films. We have undertaken to give 5000 exposures free of charge of valuable documents belonging to the Asiatic Society and also to prepare photographs of valuable non-commercial documents belonging to other societies or persons as far as practicable on payment of cost of materials.

Indian Statistical Insritute : Earmarked & spscial Enquiries Fund, Account for the Year Ending 31st March, 1944

RECEIPTS	Rs. A. P.	Rs. A. P.	PAYMENTS	Rs. A. P.	Rs. A. P.	Rs. A. P.
			Bengal Crop Survey			
			Statistical Analysis (1943-44, 44-45)			
To Block Grant for 1943-44	3,80,500 0 0		By Salary & Allowance	1,05,559 14 0		
" Advance for 1944-45	30,000 0 0		" Other expenses	53,878 2 0		
		4,10,500 0 0			1,59,438 0 0	
			Field Branch (1943-44)			
			By Salary & Allowance	1,75,060 10 0		
			" Other expenses	48,151 11 3		
					2,23,212 5 3	
			Field Branch Other expenses (1944-45)		1,768 5 0	
			By Closing balance		26,081 5 9	
						4,10,500 0 0
			Burdwan Hooghly Howrah Flushing Scheme			
To Opening balance (1942-43)	5,217 4 0		By Salary and allowance	5,960 5 0		
" Block Grant for 1943-44	10,000 0 0		" Other expenses	2,745 11 0		
		15,217 4 0			8,706 0 0	
			" Closing balance		6,511 4 0	
						15,217 4 0
			I. C. A. R. Regional Suvrey of Sugarcane Borer			
To Opening balance	3,962 10 0		By Salary & allowance	11,758 5 0		
" Block Grant	5,000 0 0		" Other expenses	5,416 11 0		
" Advance from Institute	8,212 6 0				17,175 0 0	17,175 0 0
		17,175 0 0				
			U. P. Wheat and Sugarcane Scheme			
To Opening balance		5,806 5 3	By Salary & allowance	402 9 0		
			" Other expenses	185 7 0		
					588 0 0	
			" Closing balance		5,218 5 3	
						5,806 5 3
			Bihar Crop Survey			
			Statistical Analysis			
To Block Grant		56,319 5 0	By Salary & allowance	23,078 2 0		
			" Other expenses	13,131 6 0		
					36,209 8 0	
			" Field Branch (Bengal staff on loan)		9,052 2 3	
			" Closing Balance		11,057 10 9	
						56,319 5 0
			Abstract Cash Account			
			Closing balance			
Opening balance	15,003 6 6		Earmarked and Special Enquiries Fund		48,868 9 9	
Earmarked and Special Enquiries Fund	15,000 6 6		General Fund		35,290 5 1	
General Fund	23,971 14 0					84,158 14 10
		39,975 4 6				

Director's Contribution Fund

	Donations in cash	In lieu of Prof. P. C. Mahalanobis' honorary Services	Total
		Rs. A. P.	Rs. A. P.
1932-33	189 0 0		189 0 0
1933-34	600 0 0		600 0 0
1934-35	600 0 0		600 0 0
1936-37	3,482 11 9		3,482 11 9
1939-40		2,000 0 0	2,000 0 0
1940-41		1,200 0 0	1,200 0 0
1942-43	727 5 0	11,000 0 0	11,727 5 0
1943-44		20,166 11 0	20,166 11 0
Total	5,599 0 9	34,366 11 0	39,965 11 9

6, Hastings Street Calcutta
The 29th April, 1944.

Examined and Found Correct.
P. C. Nandi & Co., *Auditors.*
Charted Accountants.
Registered Accountants.

Indian Statistical Institute : Receipts and Payments Account for the year ending 31st March, 1944

RECEIPTS

	Rs. A. P.	Rs. A. P.
To Opening Balance		38,975 4 6
,, Membership fees		780 0 0
,, Donations		191 0 0
,, Interest on Investments		110 15 0
,, Training fees and other receipts		850 1 7
,, Block Grants		
Government of india		
Depart of Education Health and Lands		45,250 0 0
Earmarked for Sankhya		2,000 0 0
Imperial Council of Agricultural Research		8,000 0 0
I.C.A.R. Regional Survey of Sugarcane Borers		5,000 0 0
Government of Bengal		
Bengal Crop Survey (1943-44)	3,80,500 0 0	
Bengal Crop Survey (1944-45)	30,000 0 0	4,10,500 0 0
B-H-H. Crop Cutting (1944-45)	10,000 0 0	
River Research Institute ..	1,555 9 0	11,555 9 0
Government of Bihar Bihar Crop Survey : 1943-44 ..		53,319 5 0
Government of Central Provinces Nagpur Labour Wnquiry		1,000 0 0
,, Contribution in lieu of Prof. P. C. Mahalanobis's services :		
Bengal Government ..	14,500 0 0(a)	
Bihar Government ..	5,666 11 0(b)	20,166 11 0
,, Provident Fund Subscriptions		3,355 14 0
,, Contribution to Leave, Salary and Provident Fund		23,217 13 6
,, Income Tax realised from staff		1,061 4 9
,, Deposit a/c		450 0 0
,, Rentals realised		10,000 0 0
,, Old advance realised		10,029 7 0
,, Advance from Prof. P. C. Mahalanobis ..		5,000 0 0
		6,53,813 5 4

PAYMENTS

	Rs. A. P.	Rs. A. P.
By Salary and allowance (including research staff)	1,61,394 4 6	
,, Contribution to Leave Salary and Provident Fund	25,823 1 6	
,, Salary and allocance (Bengal Crop Survey, Field Branch 1943-44) ..	1,75,060 10 0	3,62,278 0 0
,, Books and Journals	..	5,015 7 9
,, Laboratory Equipments ..	..	6,445 9 3
,, Establishment and Travelling ..	..	16,213 1 0
,, Postage, Printing and Contingent charges	..	11,401 2 3
,, Stationery and Store purchase ..	..	14,696 7 9
,, Other expenses (hire charge of machines, furniture and rent etc.) ..	..	28,734 1 0
,, Indian Statistical Conference ..	..	892 12 9
,, Evacuation and Emergency arrangements	..	3,239 4 0
,, Auditors' Fee	..	500 0 0
,, Repayment of old advance to Prof. P. C. Mahalanobis	..	10,005 11 6
,, Contribution to Statistical Publishing Society	..	2,870 0 0
,, Grant for Sankhya	..	2,000 0 0
,, Examination fee	..	630 0 0
,, Jute and Aus Survey (1942-43) refund	..	17 3 3
,, Bengal Government Crop Survey (1943-44) Other field expenses ..	..	48,151 11 3
,, Bengal Government Crop Survey 1944-45 Field expenses	..	1,768 5 0
,, Bihar Crop Survey (Field) ..	..	9,052 2 3
,, Construction, maintenance & repairs etc.	..	20,076 7 9
,, Giridih Housing Scheme	..	4,472 12 6
,, Suspense Account		21,179 3 3
,, Library deposit refunded ..	..	15 0 0
,, Closing balance		
Defence loan	35,122 1 0	
Cash in hand	169 1 6	
Cash in bank	48,867 12 4	84,158 14 10
		6,53,813 5 4

(a) Bengal Crop Survey : Rs. 12,000/- (Dept. of Civil Supplies No. 3995(a) dt. 21-7-43).
B-H-H-Crop Cutting : Rs. 2,500/- (Dept. of Communication & Irrigation No. 10/1 dt. 3-1-44).
(b) Bihar Crop Survey : Rs. 5,666/11/- (Revenue Dept., Bihar, No. 16301 C-29/44 dt. 19-4-44).

6 Hastings Street, Calcutta,
The 27th April, 1944.

Examined and Found Correct.
P. C. Nandi & Co., Auditors,
Registered Accountants.

APPENDIX 1. LIST OF PAPERS PUBLISHED AND REPORTS SUBMITTED IN 1943-44

1. Bhattacharya, A. On Some Sets of Sufficient Conditions leading to the Normal Bivariate Distribution. (*Sankhya*, 6(4) 1944).
2. ——— A Note on the Distribution of Sum of χ^2's (*Sankhya*, 6(4) 1944).
3. Bose, Chameli. Note on the Sampling Error in the Method of Double Sampling. (*Sankhya*, 6(3) 1943).
4. Bose, Purnendu. On Confluent Hypergeometric Series. (*Sankhya*, 6(4) 1944).
5. Bose, Purnendu and Rao, S. R. On the limiting forms of statistical distributions. (*Science and Culture*, 9(9) 1944).
6. Bose, R. C. The problem of balancing in symmetrical factorial design. (*Ind. Sc. Congress*. 1944).
7. ——— Fundamental theorem of linear estimation. (*Ind. Sc. Congress*. 1944).
8. Ghosh, Birendranath. On some artificial fields and their correlation Function. (*Science and Culture*, 8(10) 1943).
9. ——— On random distances between two rectangles (*Science and Culture*, 8(11) 1943).
10. ——— On two types of correlation function (*Science and Culture*, 9(1) 1943).
11. ——— On sampling in unknown fields (*Science and Culture*, 9(3) 1943).
12. ——— On the construction of some natural fields (*Science and Culture*, 9(5) 1943).
13. ——— Some studies in Mahalanobis' Correlation Function (*Ind. Sc. Congress*. 1944).
14. Mahalanobis, P. C. An Enquiry into the Prevalence of Drinking Tea among Middle class Indian Families in Calcutta (*Sankhya*, 6(3) 1943).
15. ——— Multi-Stage Sampling (*Ind. Sc. Congress*. 1944).
16. ——— Report on the Regional Survey of Borer Pests of Sugarcane. Parts 1—9 (*Submitted to Imperial Council of Agricultural Research*).
17. ——— Final Report on the Sample Census of the Area under Jute and Aus Paddy in Bengal 1943. (*Submitted to the Government of Bengal*).
18. ——— Preliminary Report on the Sample Census of the Area under Aman Paddy in Bengal 1943. (*Submitted to the Government of Bengal*).
19. Mahalanobis, P. C. and Ghosh, Birendranath. Statistical analysis of data relating to incidence of pests and diseases on different varieties of Sugarcane (*Indian Statistical Conference*. 1943).
20. Mathew, N. T. Lines on the Human Palm. (*Sankhya*, 6(3) 1943).
21. Nair, K. R. Certain Inequality Relationships among the Combinational Parameters of Incomplete Block Designs (*Sankhya*, 6(3) 1943).
22. ——— A Note on the problem of K-Samples (*Current Science*, 12 1943).
23. ——— The Recovery of Inter-block Information in Incomplete Block Designs (*Sankhya*, 6(4) 1944).
24. Nair, K. R. and Banerjea, K. S. A note on the fitting of straight lines when both Variables are subject to error. (*Sankhya*, 6(3) 1943).
25. Nandi, H. K. Note on the Enumeration of Non-isomorphic Solutions of Balanced Incomplete Block Designs (*Ind. Sc. Congress*. 1944).
26. Rao, C. R. Certain experimental arrangements in quasi-Latin Squares. (*Current Science*, 12(12)1943)
27. ——— Generalised Variance of Populations (*Ind. Sc. Congress*. 1944).
28. Roy, S. N. Multivariate analysis of variance when the number of populations is less than or equal to the number of characters. (*Ind. Sc. Congress*. 1944).
29. Roy, S. N. and Bose, Purnendu. Bernoulli's theorem and its analogue deduced from Tshebycheff's theorem (*Ind. Sc. Congress*. 1944).
30. Sen Gupta, J. M. A note on adjustments for first and second moments in a grouped frequency distribution split up into sub-sections (*Sankhya*, 6(4) 1944).

APPENDIX 2. OFFICERS ON DEPUTATION FOR TRAINING AND SCIENTIFIC CONSULTATION : 1943-44

1. S. M. Sarkar, Ph. D. (*Lecturer in Botany, Calcutta University*) February, 1943.
2. S. P. Bose, B.Sc. (Wales). (*Institute of Rural Reconstruction, Sriniketan, Birbhum*) April, 1943.
3. Md. Abdur Rahaman, M.Sc. (*River Research Institute, Bengal*) May, 1943; Feb. and March, 1944.
4. I. B. Chatterjee, M.Sc..(Ag.), L.Ag. (*Physiological Chemist, Government of Bengal, Dacca*) May, 1943.
5. P. R. Ray, M.Sc. (*Statistician, Department of Agriculture, Bengal*) May, 1943 and March 1944.
6. K. P. Pillai, M.Sc. (*Physiological Assistant, Ravenshaw College, Cuttack*) June 1943 ; Jan. Feb., March 1944.
7. A. K. Chakravarty, M.Sc. (*Rice Research Officer, Bankura*) June 1943.
8. S. Hedayetullah, Ph.D. (Cantab) (*Economic Botanist, Department of Agriculture, Bengal*) June 1943.
9. P. B. Bhattacherjee, M.Sc. (*St. Stephen's College, Delhi*) 21 June to 19 July 1943.
10. J. B. Mathur, Dip. Hort. (London, Wye) (*Horticulturist, Orissa*) 2 July to 2 October, 1943.
11. Subrata Sen, M.Sc. (*Department of Agriculture, Bengal*) 14 July to 4 August and 1 to 9 October 1943.
12. S. C. Chakravarty (*Rice Research Officer, Chinsura*) 21 July 1943 and 18 August 1943 and 14 February 1944.
13. P. G. Chowdhury, M.B., D.P H., (*Supdt. of Vital Statistics, Public Health Dept. Bengal*) July 1943.
14. J. R. Pal, B.Ag. (*Assistant Rice Research Officer, Bankura*) 29 July to 6 August 1943 and 11 September to 15 September 1943 and 18 March 1944.
15. M. C. Chakravarty, M.Sc. (*Lecturer in Statistics, Dacca University*) 28 September to 27 October 1943.
16. Suprokash Sen, M.Sc. (*Imperial Veterinary Research Institute, Mukteswar, U. P.*) October, 1943.
17. Balakrishna Nayar, M.Sc. (*Assistant Marketing Officer, Government of India*) December 1943.
18. M. C. Tewari. (*Imperial Veterinary Research Institute, Mukteswar, U. P.*) 12 October, 1943 to 12 February, 1944.
19. Satyabrata Sen, B.Sc. (Edin,) (*Statistician, Indian Tea Market Expansion Board*) February, and April 1944.
20. N. C. Sen Gupta, D.Sc. (*Burmah Oil Company (India) Ltd.*) Working from March, 1944

APPENDIX 3. STUDENTS IN THE TRAINING SECTION AND TECHNICAL APPRENTICES: 1943-44

1. Miss Amita Choudhury, M.A. (*Dacca*) April 1943—June 1943.
2. Md. Fashiuddin, M.A. (*Aligarh*) July 1943—(*Still working*).
3. S. Janardan Poti, M.A. (*Benares*) July 1943—January, 1944. (*Institute staff.*)
4. T. V. Rangarajan. B.A. (*Madras*) July, 1943—August 1943.
5. K. Venkata Krishnan, B.A. (*Hons.*) (*Madras*) August 1943—November 1943.
6. Salil Kumar Ray Chowdhury, M.A. (*Calcutta*) September 1943—October 1943.
7. Sushil Kumar Banerjee, M.A. (*Calcutta*) February 1943—January, 1944.
8. B. Harinarayana. M.Sc. (*Calcutta*) March 1943—October 1943.
9. N. R. Ramasubramanian, B.Sc. (*Madras*) March 1943—December 1943.
10. V. Ramachandra Rao, M.A. (*Andhra*) March 1943—January 1944.
11. Manindra Nath Ghosh, M.A. (*Calcutta*) April 1943—November 1943.
12. Nirmal Kumar Chakravarty, M.A. (*Calcutta*) September 1943—
13. Harikinkar Nandy. M.Sc. (*Calcutta*) September 1943—
14. K. C. Cheriyan, M.A., M.Sc. (*Madras*) September 1943—
15. R. P. Bhargava, M.Sc. (*Allahabad*), M.Sc. (*Calcutta*) November 1943—
16. C. Radhadkrishna Rao, M.A. (*Andhra*), M. A. (*Calcutta*) November, 1943—
17. Ram Krishna Mookherjee, M.Sc. (*Calcutta*) January, 1944—
18. Gouri Sankar Chatterjee, M.A. (*Calcutta*) February, 1944—

Kumar Gayen, M.A. (*Calcutta*) February, 1944.—

INDIAN STATISTICAL INSTITUTE

PROCEEDINGS OF THE COUNCIL OF THE INDIAN STATISTICAL INSTITUTE DATED 29th APRIL, 1944.

A meeting of the Council of the Indian Statistical Institute was held at 3 P. M. on Saturday the 29th April 1944 in the Statistical Laboratory, Presidency College, Calcutta.

The Hony. Secretary stated that a telephone message was received from the house of Mr. N. R. Sarker, President, stating that Mr. Sarker had become suddenly indisposed and regretted his inability to attend the meeting.

Present : Mr. B. B. Das Gupta was voted to the Chair (in the absence of the President). Prof. K. B. Madhava, Mrs. Chameli Bose, Messrs. Raj Chandra Bose, Samarendranath Roy, Prof. Khagendra nath Chakravarti, Dr. Harish Chandra Sinha, Prof. S. N. Bose, Prof. P. C. Mahalanobis (Hony. Secretary) and Mr. Haricharan Ghosh.

1. Proceedings of the meeting of the Council dated 30th September, 1943 ; 4th January, 9th February, and the 21st February 1944 were read and confirmed.

2. *Election of Members* : The following gentlemen were elected ordinary members of the Indian Statistical Institute for the year 1944-45.

> Mr. Sydney John William Wells, "Nev-Del", 29, Kensington Street, Haberfield, N. S. W., Australia ; Mr. Debidas Mazumdar, Calcutta ; Kumar Bimal Chandra Sinha, Calcutta ; Krishnadas Mansingh Lalla, Digboi. Assam.

3. *Bengal Crop Survey Scheme* 1944-45 : Considered item No.I(iii) of the Finance Committee dated 27th April 1944. It was reported that the scheme for a sample survey of crops in Bengal covering both the summer-monsoon and the winter seasons in 1944-45 at an estimated cost of Rs. 4,20,000/- (inclusive of dearness allowance at 1943 rates which however would be modified in case any changes were made in the scale) had been approved by the Government of Bengal. The Hon. Secretary mentioned that the Government of Bengal had agreed in No. 2071 D.C.S. dated Calcutta the 17th March 1944 to place the part-time services of Mr. N. C. Chakravarti, B.C.S., at the disposal of the Institute as part-time Superintendent of Statistical Surveys especially for the Field Branch and had permitted him to receive a fee of Rs. 150/- per month out of Institute funds. The Hony. Secretary also stated that certain changes in the number and pay of the field staff had been considered necessary by the Superintendent of Statistical Surveys and certain modifications were proposed to be made within the limits of the grant.

Resolved that the Hon. Secretary be requested to place the revised budget before the Finance Committee in due course, but in the meantime he be authorised to make all necessary arrangements for the Bengal Crop Survey Scheme.

4. *Bihar Crop Survey Scheme* : Considered item No.I(ii) of the Finance Committee dated 27th April 1944. In continuation of earlier reports the Hon. Secretary stated that the preparatory work for the Bihar Crop Survey was completed early in February. He sent 72 of the Bengal field staff on loan to Bihar and 108 men were recruited locally by the Government of Bihar. The field survey started on the 12th February 1944, and was closed on the 13th April. The Hon. Secretary went to Patna on the 23rd March, and had a long interview with H. E. Mr. Mudie, Governor of Bihar, on the 24th and other officials. He then went to inspect the field work on the 24th and 25th at Bikramgunj in Shahabad district where conferences with statistical and field workers were arranged. The exploratory survey was carried out with such great success that in consultation with Rai Bahadur S. P. Sinha, Superintendent of Statistical Surveys, it was decided to organize a full scale provincial survey in the next *bhadoi* season thus stepping up the original programme by 16 months. A definite scheme at an estimated cost of six lakhs of rupees (including Rs. Rs. 2,48,000 for statistical work to be given as a block grant to the Institute) had been submitted to the Government of Bihar. An application for acquistion of land at Giridih had been submitted and it was

believed that necessary action was being taken on it by Government. Because of the stepping up of the programme, 200 or more workers would have to be accommodated at Giridih ; pending acquisition of land it might be necessary to put up additional rooms in houses rented by the Institute.

Resolved that the recommendation of the Finance Committee be accepted and the Hon. Secretary be authorized to take all necessary action in the Bihar Crop Survey Scheme and that the proposal to increase accommodation in rented buildings on the lines suggested by the Hon. Secretary be approved in principle and he be requested to submit a detailed scheme to the Finance Committee before actual construction work is started.

5. *Journal Committee* : Resolved that the Journal Committee be reconstituted as follows for one year with powers to co-opt :—

Prof. S. N. Bose, Prof. K. B. Madhava, Dr. H. Sinha, The Hon'ble Mr. Justice T. J. Y. Roxburgh. Dr. P. V. Sukhatme, Mr. R. C. Bose (Secretary) and Prof. P. C. Mahalanobis (Editor).

6. *Finance Committee* : Resolved that the Finance Committee be reconstituted as follows for one year with powers to co-opt :—

The President, Treasurer, & Secretary, Prof. K. B. Madhava and Prof. K N. Chakravarti (Joint Secretaries), Prof. H. C. Ghosh Mr. B. B. Das Gupta, Mr. N. C. Chakravarti, Mr. Satyendranath Ray, Mr. Samarendranath Roy. (Three to form quorum).

7. *Increase in Membership fees and Subscriptions for Sankhya* : Considered the question of increasing the subscription for Sankhya and the subscription of members (who are supplied Sankhya free) on account of the abnormal rise in the price of paper, stationery and printing.

Resolved that an increase in the subscription for Sankhya from Rs. 20/- to Rs. 30/- per volume of, 4 parts and Rs. 9/- per single part be approved, and also resolved that the Council recommends to the General Meeting of the Institute that the subscription for members be increased from Rs. 15/- to Rs. 20/- per year and the life-membership fee from Rs. 150/- to Rs. 250/-.

8. *Provident Fund* : Considered item No.VI(3) of the Finance Committee dated 27th April 1944 Considered the changes in the Provident Fund Rules and other recommendations made by Prof. K. B. Madhava which had been approved generally by the Finance Committee.

Resolved that the following changes in the Rules of the Provident Fund be approved and be submitted to the General Meeting of the Council for confirmation. Also resolved that the Finance Committee be requested to make necessary arrangements for the maintenance of the ledger and supervise the general working of the Provident Fund.

Changes in the Rules of the Provident Fund

Rule No.	As now Existing	Proposed Modification	Brief statement of reason for change.
7(iii)	"Before the 10th day of every month"	"At the end of Financial year"	To avoid unnecessary adjustment on account of members who withdraw in the course of the year, and also to render the work more practical
	"on the same or next working day"	"From time to time"	It is not practicable to do as now required in the Rules and is not in practice being done.
9(i)	"Board of Trustees"	"Council of the Institute"	Associations such as Calcutta University Provident Fund, Indian Association for the Cultivation of Science, Asiatic Society of Bengal, retain the final authority and vesting of property in their respective Councils of Management and not in Trustees. There are in these bodies no Trustees.
	"All decisions of the Board of Trustees"	"their decisions"	
(i)	"Trustees at their disretion"	"Council of the Institute at their discretion on the advice of the Trustees"	

10.	At the beginning of the Rule	Add the words "with the approval of the Council of the Institute."	
19.	At the end of the Rule	Add the words "which will be effective after the approval of the Council of the Institute is given thereto"	
18.	At the end of the first sentence	Add the sentence "Any surplus remaining in the Provident Fund out of the Employer's contributions forfeited to the Fund shall vest in the Fund, and the Council of the Institute shall from time to time determine and dispose of such surplus"	This was an omission in the existing Rules.
14(i)	Under item "Period of service : (1), (2) & (3)	Substitute the word "Three years or less......Nil" and renumber (4) to (8) as (2) to (6).	The change will not affect present members. It will only affect future entrants and the modification is intended to induce them to serve sufficiently long.

9. *Population Data Committee* : Considered item No.I(iv) of the Finance Committee dated 27th April 1944. At the request of the Government of India the Hon. Secretary had gone to Delhi at the end of March for a joint discussion with Mr. S. H. Oulsnam, I.C.S. (Joint Secretary, Department of Education, Health and Lands), Lt. Col. E. Cotter, I.M.S. (Public Health Commissioner), and Mr. M. W. M. Yeatts, I.C.S. who had acted as Census Commissioner in 1941. As a result of this discussion the Government of India, was to appoint a Population Data Committee with a view to examine and advise Government on data relating to the growth of population in India. The Hon. Secretary had suggested that the statistical work in the present connexion be done in the Institute for which a suitable block grant would be sanctioned by the Government of India, and had also made certain suggestions about associating Prof. Madhava with the work.

Resolved that the recommendation of the Finance Committee be accepted and the action taken by the Hon. Secretary in regard to the Population Data Committee be approved, and he be authorized to make all necessary arrangements in this connexion in anticipation of sanction by the Council.

10. *Superintending Statistician* : Considered item No.I(v) of the Finance Committee dated 27th April 1944.

Resolved that the proposal of the Hon. Secretary be approved and that the senior workers in charge of Giridih, Calcutta and Baranagore Branches of the Laboratory be called Superintending Statisticians, and that Rs. 300/- per month be sanctioned for special allowance of these posts.

11. *Bursars* : Considered item No.I(vi) of the Finance Committee dated 27th April 1944. For more than two years Prof. Haricharan Ghosh had ungrudgingly undertaken the heavy and onerous task of scrutinizing and passing all bills for payment. The volume of work was however increasing appreciably and the Hon. Secretary suggested that arrangements should be made to have the work done by one or two part-time Bursars for which a part-time allowance might be sanctioned.

Resolved that the recommendation of the Finance Committee be accepted and Rs. 200/- per month be sanctioned for payment of part-time allowance for the scrutiny and payment of bills.

The Council places on record its deep appreciation of the services rendered by Prof. Haricharan Ghosh for more than two years in scrutinizing and passing all bills in a purely honorary capacity.

Resolved that Prof. Haricharan Ghosh and Prof. Khagendranath Chakravarti be requested to act as Bursars of the Institute for 1944-45 and they be offered a part-time honorarium of Rs. 100/- per month for this work and the Secretary be requested to take necessary action in this matter.

12. *Visiting Professors* : Considered item No. I(vii) of the Finance Committee dated 27th April, 1944. Prof. R. A. Fisher had been invited to come to India as Visiting Professor in 1937-38 and Prof. H. Hotelling in 1939-40, but the programme was kept in abeyance under war conditions. As a result of informal discussions with Prof. A. V. Hill, F.R.S., Secretary of the Royal Society, London, and Government of India officials the Hon. Secretary thought that it would be desirable to revive the practice.

Resolved that the proposal be approved and Hon. Secretary be authorized to take necessary action.

13. On the recommendation of the Finance Committee dated 27th April 1944 resolved that a security account with the Imperial Bank of India, Calcutta, be opened with securities such as Defence Loans for the purpose of taking overdraft from the current account 'B' of the Indian Statistical Institute with them to the extent of Rs. 25,000/- (Rupees twentyfive thousand only). The Security Account will also be operated by the Hon. Secretary, Prof. P. C. Mahalanobis. He shall have also the power to withdraw, transfer or negotiate such securities as are being deposited in the above mentioned account from time to time.

14. *Rules for the Acceptance of Block Grants* : Considered item No.I(viii) of the Finance Committee dated the 27th April 1944.

Resolved that the attached rules for the acceptance of block grants to the Institute be approved and be brought into immediate effect and that all authorities proposing to make project grants for the year 1944-45 be informed accordingly.

RULES FOR ACCEPTANCE OF BLOCK GRANTS

(1) Block Grants for projects or specific enquiries would be accepted by the Indian Statistical Institute on the clear understanding that the actual work would be done on a unitary basis by the staff of the Institute.

(2) The accounts of the Institute would be audited by a firm of Chartered Accountants and a statement of accounts certified by the auditors and endorsed by the Hony. Secretary would be submitted to each authority making a block grant and should be acceptable as fulfilling the conditions of the grant in regard to accounts.

(3) Separate accounts and vouchers would be maintained for exclusive charges relating to particular projects as far as practicable but all charges for which such exclusive accounts and vouchers cannot be kept will be allocated to different heads by appropriate authorities of the Institute in consultation with the auditors of the Institute and the decision of the Council of the Indian Statistical Institute in this regard shall be final.

(4) Since all projects and specific enquiries are undertaken as an integral part of the work of the Institute all stock purchased such as maps, stationery, furniture etc. would remain the general stock of the Institute and charges on this account would be allocated to different projects or grants as provided in rule(3).

(5) Travelling expenses and other incidental charges would be supported by exclusive vouchers wherever practicable but where journeys are undertaken for the benefit of more than one project suitable allocation of the charge would be made by the appropriate Institute authorities at their discretion.

(6) As the work of the Institute is carried out on a unitary basis all employees are engaged as Institute staff without specific allocation to particular projects ; and as the staff have to be maintained and leave granted even after the conclusion of particular projects on which they might be working from time to time a consolidated charge of 16¼ per cent of the amounts paid out as salaries to staff be deducted on account of leave and provident fund contribution and credited to the Reserve Funds of the Institute.

(7) In undertaking any project work certain services and facilities in the way of furniture, machines library and stocks have to be placed at the disposal of each project work as necessary, but detailed allocation of expenditure for such overhead items being not possible a calculated sum to cover depreciation and use of machinery, use of library, furniture and value of stock used should be recovered from the grants by recovering pro rata according to the amounts of the block grants.

(8) A budget estimate prepared on the above basis shall be furnished to the authority proposing to make any project grant at the time of acceptance of the work but the Finance Committee and the Council of the Institute shall have full power to make necessary reappropriations within the total grant. As the Institute is not a profit-making association the balance, if any, of any particular grant shall be returned after the accounts of the same are finally adjusted and passed by the Council of the Institute.

(9) In view of the fact that Prof. P. C. Mahalanobis, Secretary of the Institute, has expressed a desire of looking after projects work in a purely honorary capacity for the present, certain authorities making project grants such as the Government of Bengal and the Government of Bihar have from time to time sanctioned contributions to the Institute in lieu of the services of Prof. Mahalanobis. The Council of the

Institute have decided that all such contributions would be credited to a fund to be called the "Director's Contribution Fund" which would for the present be kept invested in Government securities. This fund would be utilised in consultation with Prof. P. C. Mahalanobis for such purposes as the Council may think proper.

15. *Director's Contribution Fund* : Considered item No.VII(i) of the Finance Committee dated 27th April 1944. In accordance with resolution No.3(c) of the Council dated 21st February 1943 considered the following statement relating to the donations received from Prof. and Mrs. P. C. Mahalanobis and contributions received in cash from different authorities in lieu of the services rendered by Prof. P. C. Mahalanobis in connexion with different projects.

	Donations in cash			In lieu of Prof. Mahalanobis's honorary services			Total		
	Rs.	A.	P.	Rs.	A.	P.	Rs.	A.	P.
1932-33	189	0	0				189	0	0
1933-34	600	0	0				600	0	0
1934-35	600	0	0				600	0	0
1936-37	3,482	11	9				3,482	11	9
1939-40				2,000	0	0	2,000	0	0
1940-41				1,200	0	0	1,200	0	0
1942-43	727	5	0	11,000	0	0	11,727	5	0
1943-44				20,166	11	0	20,166	11	0
Total	5,599	0	9	34,366	11	0	39,965	11	9

Resolved that the above statements be published as an appendix to the accounts. On the recommendation of the Finance Committee also resolved that (a) the whole of the above amount be invested in Government securities for the present ; (b) a separate account to be called the "Director's Contribution Fund" be maintained for the above fund ; and (c) in accordance with the resolution of the Council dated 21st February 1944 expenditure out of this fund be incurred in consultation with Prof. P. C. Mahalanobis.

16. *Capital Fund* : Considered item No.VII(2) of the Finance Committee dated 27th April 1944.

Resolved that the recommendation of the Finance Committee be accepted and (a) receipts on account of donations, membership fee, leave and provident fund contributions and sundry receipts of the nature of capital grants (other than additions to the Director's contribution Fund) be treated as a capital fund ; (b) a separate account be maintained for the same in such form as may be considered suitable by the Auditors of the Institute; and (c) that advances made to the General Fund of the Institute out of the above receipts be restored at the earliest opportunity and the total amount be invested in Government securities as early as possible.

17. *Budget Estimate for* 1944-45 : Considered the draft budget estimate for the year 1944-45 as prepared by the Finance Committee.

Resolved that the attached budget estimate for the year 1944-45 as modified by discussions in the Council be passed.

18. The Hon. Secretary placed before the meeting the draft Annual Report.

Resolved that the draft Annual Report be forwarded to the Annual General Meeting.

19. The Hon. Secretary placed before the Council the Audited Accounts for 1943-44 which had been considered by the Finance Committee on the 27th April 1944 and also statements relating to project and other funds.

Resolved that the Audited Accounts and statements relating to funds be forwarded to the Annual General Meeting.

20. *Institute Examinations* : (*i*) To suit the convenience of the candidates and the Board of Examiners, resolved that the Institute Examinations be postponed till July 1944 and the Hon. Secretary be authorized to fix the dates.

(ii) Considered a letter dated 1st March 1944 from Mr. P. Satagopan, Labour Department, Simla, suggesting that since there is a Branch of the Institute at Delhi arrangements be made for holding Institute Examinations in Statistics at Delhi.

Resolved that the proposal to hold Institute examinations in Delhi and outside centres be approved in principle provided necessary facilities are available and that the Hon. Secretary be authorized to make such arrangements in this matter as may be found possible.

21. *Indian Science Congress Association* : Read a letter dated 17th March 1944 from the General Secretary, Indian Scinece Congress Association requesting that delegates be appointed to represent the Institute at the next session of the Indian Science Congress to be held at Nagpur in January 1945.

Resolved that the Hon. Secretary be requested to select the delegates in consultation with the President at the proper time.

22. *Contribution* : The Hon. Secretary placed before the meeting G.O. No.1630R dated 19th April 1944 from the Government of Bihar sanctioning a contribution of Rs. 15,000/- annually to the Indian Statistical Institute with effect from 15th November 1943 until further orders in lieu of the services of Prof. P. C. Mahalanobis in connexion with the Bihar Crop Survey.

Recorded

23. *Resignation of Asst. Secretary* : Resolved that the resignation of Dr. Atindranath Bose from the post of Assistant Secretary be accepted with effect from the date of expiry of his leave.

24. *Appointment of Asst. Secretary* : On the recommendation of the Finance Committee resolved that the appointment of Mr. M. K. Dutt (a Burma Government servant on surplus leave on half pay) as Assistant Secretary be extended for a further period of one year on the present terms with effect from 1st April 1944.

25. *Leave* : Considered the application dated 25th March 1944 from Mr. K. Raghavan Nair for extension of leave till 30th June 1944.

Resolved in modification of Council Resolution No. 7 dated 9th February 1944 that Mr. K. Raghavan Nair be granted leave without pay till 30th June 1944.

Indian Statistical Institute : Budget Estimate for the 1944-45

RECEIPTS				
Opening Balance		49,036	13	10
Grant from the Governemnt of India		51,000	0	0
Grant from the Government Of India for Sankhya		2,000	0	0
Membership Subscriptions		1,000	0	0
Miscellaneous Receipts		1,263	2	2
Goverment of Bengal :				
Bengal Crop Survey		4,07,500	0	0
B. H. H. Flushing Scheme		10,000	0	0
Government of Bihar :				
Bihar Crop Survey		2,48,200	0	0
Contributions in lieu of Prof. P. C. Mahalanobis' Service				
Bengal Crop Survey	12,500			
B.H.H. Crop Cutting	2,500			
Bihar Crop Survey	15,000	30,000	0	0
		8,00,000	0	0

EXPENDITURE			
Salary of Statistical & Research Staff	2,83,000	0	0
Director's Contributions Fund (as per contra)	30,000	0	0
Establishment & Travelling	20,000	0	0
Books & Journals	12,000	0	0
Laboratory Equipments	7,500	0	0
Bengal Crop Survey (Field Branch)	2,58,000	0	0
Stationery & Stores	23,000	0	0
Postage, Printing & Contingen.	15,000	0	0
Other Expenses	30,000	0	0
Indian Statistical Conference	1,000	0	0
Evacuation & Emergency Arrangements	4000	0	0
Grants for special activities at Giridih Office	5,000	0	0
Supernumerary Post etc.	3,000	0	0
Grant for Sankhya	2,000	0	0
Contributions to Sankhya	1,500	0	0
Giridih Housing Scheme	30,000	0	0
Closing Balance in Bank and Cash	45,000	0	0
Closing Balance in Government Securities	30,000	0	0
	8,00,000	0	0

B. DAS GUPTA
Chairman.

P. C. MAHALANOBIS
Hony. Secretary.

INDIAN STATISTICAL INSTITUTE

PROCEEDINGS OF THE ANNUAL GENERAL MEETING

DATED 29th APRIL, 1944.

The Annual General Meeting of the Indian Statistical Institute was held at 4 p.m. on Saturday the 29th April, 1944 in the Statistical Laboratory, Presidency College, Calcutta.

The Hon. Secretary reported that a message had been received from the house of Mr. N. R. Sarker, President, stating that Mr. Sarker had become suddenly indisposed and regretted his inability to attend the Meeting.

In the absence of the President, Prof. S. N. Bose, (*Dacca University*), was voted to the chair.

The following members were present :—

Mr. H. Bhattacharji	*Statistical section, D.G.M.P's office, Dept. of Supply, Calcutta.*
Mrs. C. Bose, B.Sc.	*Statistical Laboratory, Calcutta.*
Mr. R. C. Bose, M.A.	*Lecturer, Calcutta University*
Prof. K. N. Chakravarti, M.Sc.	*Presidency College, Calcutta.*
Mr. N. C. Chakravarti, M.A., B.C.S.	*Special Officer, Revenue Department, Bengal.*
Mr. B. B. Das Gupta, M.A., B.C.S.	*Special Officer, Finance Department, Bengal*
Prof. H. C. Ghosh, M.A., P.R.S.	*Calcutta University.*
Prof. K. B. Madhava, M.A., A.I.A.	*Mysore University.*
Mr. S. N. Ray, I.C.S.	*Financial Adviser, Department of Civil Supplies, Bengal.*
Mr. S. N. Roy, M.Sc.	*Lecturer, Calcutta University.*
Mr. Sadasiv Sen Gupta, M.Sc., B.Sc. (Lond.)	*Statistical Officer, E. I. Rly.*
Dr. H. C. Sinha, M.Sc., Ph.D.	*Calcutta University.*
Prf. P. C. Mahalanobis	*Honorary Secretary.*

1. *Annual Report and Audited Accounts :* The Honorary Secretary placed before the meeting the Annual Repopt and Audited Accounts for 1943-44 as approved by the Council and mentioned that the Accounts had been audited by Messrs. P. C. Nandi & Co., Chartered Accountants.

Resolved that the Annual Report and Audited Accounts (*annexed*) for 1943-44 be adopted.

Proposed by Mr. B. B. Das Gupta
Seconded by Mr. Haricharan Ghosh (*Carried unanimously*).

2. *Election af Office-Bearers and Members of the Council :* The Honorary Secretary reported that in accordance with Bye-law 3 for election of office-bears and members of the Council, nomihations for 1944-45 approved by the Council of the Institute had been circulated together with the notice of the Annual General Meeting on the 6th April, 1944. As the number of nominations was in no case in excess of the number of vacancies, no voting was necessary and no voting paper was therefore issued.

The election of the following office-bearers and members of the Council as circulated was then annocnced from the chair.

President

Mr. N. R. Sarker,	*Calcutta.*

Vice-Presidents

Dr. P. N. Banerjea, M.A., D.Sc., M.L.A. (Central),	*Calcutta*
Mr. E. T. Cordwell,	*Messrs. Bird & Co., Calcutta.*
Sir C. D. Deshmukh, C.I.E.,	*Governor, Reserve Bank of India.*
Sir A. H. Ghuznavi, Kt., M.L.A. (Central),	*Messrs. A. H. Ghuznavi & Co., Calcutta.*
Dr. J. B. Grant, M.D., Dr. P. H.,	*Director, All-India Institute of Hygiene and Public Health, Calcutta.*
Dr. S. P. Mookerjee, M.A., B.L., M.L.A., D.Litt., Bar-at-law,	*Calcutta.*
Mr. Chand Mull Rajgharia	*Mica Merchant, Giridih.*
Sir Shri Ram, Kt.,	*Delhi Cloth and General Mills, Delhi.*
Sir C. V. Raman, D.Sc., F.R.S., N.L.,	*Indian Institute of Science, Bangalore.*
The Hon'ble Mr. Justice T. J. Y. Roxburgh B.A., Bar-at-law, C.I.E., I.C.S.,	*Judge, Calcutta High Court.*
The Hon'ble Mr. H. S. Suhrawardy,	*Minister in charge of Civil Supplies, Government of Bengal.*

Treasurer

Dr. Satya Churn Law, M.A., B.L., Ph.D., F.Z.S., M.B.O.U.,	*Calcutta.*

Secretary

Prof. P. C. Mahalanobis,	*Statistical Laboratory Calcutta.*

Joint Secretaries

Mr. Raj Chandra Bose, M.A.,	*Lecturer, Calcutta University.*
Prof. K. N. Chakravarti, M.Sc.,	*Presidency College, Calcutta.*

Members of the Council for 1944-45 .

Mr. B. P. Adarkar, M.A.,	*University of Allahabad.*
Mrs. C. Bose, B.Sc., (London),	*Statistical Laboratory, Calcutta.*
Dr. G. S. Bose, M.D., D.Sc.,	*Head of Department of Psychology, Calcutta University.*
Prof. S. N. Bose, M.Sc.,	*University of Dacca.*
Mr. P. H. Carpenter, O.B.E., F.I.C., F.C.S.,	*Director, Indian Tea Association, Tocklai, Assam.*
Mr. Nihar Chandra Chakravarti, M.A.,	*Special Officer, Revenue Department, Bengal.*
Mr. B. B. Das Gupta, M.A.,	*Special Officer, Finance Department, Bengal.*
Prof. H. C. Ghosh, M.A., P.R.S.,	*University of Calcutta.*
Mr. J. V. Joshi, M.A., O.B.E.,	*Senior Economist, Reserve Bank of India, Bombay.*
Prof. F. W. Levi, Dr. Phil. Nat.,	*University of Calcutta.*
Prof. John Maclean,	*Wilson College, Bombay.*
Mr. A. C. Mukherji, M.A., Dip. de Statistician, F.S.S.	*Director of Statistics, Baroda State, Baroda.*
Mr. K. R. Nair, M.A., M.Sc.,	*Food Department, Government of India.*
Dr. K. C. K. E. Raja, L.R.C.P. & S., D.P.H., D.T.M. & H.,	*Deputy Public Health Commissioner with the Government of India, Delhi.*
Dr. V. K. R. V. Rao, M.A., Ph.D.,	*University of Delhi.*
Mr. S. N. Ray, B.A. (Cantab), I.C.S,	*Financial Adviser, Department of Civil Supplies, Bengal.*
Mr. B. Ranga Reddy,	*Secunderabad.*
Mr. Samarendra Nath Roy, M.Sc.,	*Lecturer, Calcutta University.*
Dr. H. C. Sinha, M.Sc., Ph.D.,	*Lecturer, Calcutta University.*
Mr. P. V. Sukhatme, Ph.D., D.Sc.,	*Statistician, Imperial Council of Agricultural Research, Delhi.*

3. *Recomuendotions from the Council :* Considered the recommendation of the Council that "in view of the abnormal rise in the price of paper and stationery and printing, the membership fee be raised to Rs. 20 per annum for ordinary members, Rs. 250 for life members, and the subscription for Sankhya to Rs. 30 per volume of four parts and Rs. 9 for each part with effect from April, 1944".

Resolved that the proposal be approved.

Proposed by Dr. Haris Chandra Sinha

Seconedd by Mr. Samarendra Nath Roy

(*Carried unanimonsly*).

4. Provident Fund Rules as recommended by the Council (printed on pp.17-18) be confirmed.

Proposed by : Prof. K. B. Madhava

Seconded by : Mr. B. B. Das Gupta

(*Carried unanimously*).

5. *Appointment of Auditors :* Resolved that Messrs. P. C. Nandi ffl Co., Chartered Acconntants be appointed auditors for the year 1944-45 on an honorarium of Rs. 600. (3 to 4 quoram)

Proposed by Mr. N. C. Chakravarti

Seconded by Mr. 8. Sen Gupta

(*Carried unanimously*).

6. *Apgointment of Committees :* Resolved that as recommended by the Council the following Committees be reconstituted as follows :

(*i*) *Journal Coumittee* (1 year): Prof. S. N. Bose, Prof, K.B. Madhava; Dr. H. Singha, The Hon'ble Mr. Justice T. J. Y. Roxburgh, Dr. P. V. Sukhatme, Mr. R. C. Bose, (*Secretary*), and Prof. P. C. Mahalanobis, (*Editor*).

(*ii*) *Finance Committee* (1 year): The President, Treasurer and Secretary, Prof. K. B. Madhava and Prof. K. N. Chakravarti (*Joint Secretarias*) Mr. B. B. Das Gupta, Mr. N. C. Chakravarti, Prof. H. C· Gho e, Mr. Satyendra Nath Ray and Mr. Samarendra Nath Roy.

Proposed by Mrs. Chameli Bose

Seconded by Mr. H. G. Bhattacharji

(*Carried unanimously*)

S. N. BOSE
Chairman

P. C. MAHALANOBIS
Honorary Secretary,

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Edited By P. C. MAHALANOBIS

VOL. 7 PART 2 1945

THE LOGICAL INVERSION OF THE NOTION OF THE RANDOM VARIABLE

BY R. A. FISHER
University of Cambridge

The mathematical concept of the random variable has been of value in giving formal precision to all statements of the theory of probability which depend on the notion of frequency. For the definition of such a random variable we require a probability P, such that $0 \leqslant P \leqslant 1$, defined as a function of the variable, in such a way that the definition may involve parameters,

$$\mathrm{P}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-\frac{1}{2}u^2}\, du \qquad \ldots (1)$$

in which the distribution of the random variable x is defined, and is dependent on the two parameters μ and σ. P is then said to be the probability that the random variable shall be less than any assigned value x.

This definition brings into clear light a point, which has often been indicated by logical writers, but which is obscured by the forms of common speech, that in respect of any concrete object or event, the value of the probability will depend on the population to which, for purposes of discussion, that object or event is appropriately deemed to belong. The probability, or frequency, with which it rains at Manchester (tacitly including equally all periods of the day) is not necessarily the same as the frequency with which it rains at 9 a.m. in Lancashire (tacitly including equally all areas in the Country) ; consequently an evaluation of probability, to be mathematically correct, must utilise all the information supplied by the data, and must explicitly exclude only those elements which in any particular problem are known to be irrelevant. Recognising this, however, we shall be free, in relation to any concrete event, to evaluate probabilities relevant to that event regarded as a member of any hypothetical population to which it belongs, and for which we possess the appropriate information.

In the observational sciences, the observations, which are particular and concrete events, seem to be capable of use, and are now freely used, as a basis for probability statements respecting parameters, the values of which, apart from the information given by such observations, are unknown. This type of reasoning arises from what have been called the tests of significance, and the probability statements arrived at are distinguished as statements of fiducial probability, to avoid confusion with the earlier usage of "inverse probability" on the basis of Bayes' postulate. The object has been to arrive at probability statements of practical use without assuming Bayes' postulate, not, of course, to find another method of calculating the probabilities a *posteriori* of Bayes, which, without that postulate, are obviously indeterminate.

In recent times one often-repeated exposition of the tests of significance, by J. Neyman, a writer not closely associated with the development of these tests, seems liable to lead mathematical readers astray, through laying down axiomatically, what is not agreed or generally true, that the level of significance must be equal to the frequency with which the hypothesis is rejected in repeated sampling of any fixed population allowed by hypothesis. This intrusive axiom, which is foreign to the reasoning on which the tests of significance were in fact based, seems to be a real bar to progress, for I am not aware that Neyman, or his followers, have succeeded in putting forward any tolerable solution even to the problem of the comparison of the means of two samples drawn from different normal populations : a satisfactory solution of which was given by Behrens in 1929. The current tables, both those of Sukhatme and my own, are based on formulae equivalent to those of Behrens though somewhat more general.

The purpose of this note is therefore to discuss, without prejudice to the question of axiomatic foundations, the process of reasoning by which we may pass, without arbitrariness or ambiguity, from forms of statement in which observations are regarded as random variables, having distribution functions involving certain fixed but unknown parameters, to forms of statement in which the observations constitute fixed data, and frequency distributions are found for the unknown parameters regarded as random variables.

In tests of significance we define a class of events (samples) having a certain specified frequency (5 per cent, 1 per cent etc., according to the level of significance chosen) among a population A, of samples to which we regard the sample observed, to which the test of significance is to be applied, to belong. This population must resemble the sample observed in all relevant respects ; indeed the argument is not impeded if it be taken to resemble the sample in all observable respects.

Thus in testing the significance of the mean of a normal sample, we have as data the observed values $x_1, x_2, \ldots, x_n$, and are concerned to draw what inference may be possible about the mean of the hypothetical normal population, from which these observations form by hypothesis a random sample. It will in this case be sufficient to consider a population of samples characterised by the two statistics $\bar{x}$ and s, and by the ancillary statistic N, defining these by the relations

$$N\bar{x} = S(x)$$
$$(N-1)s^2 = S(x-\bar{x})^2$$

The reason for which the limited specification is sufficient is that it has been shown that, for given values of N, $\bar{x}$ and s, the distribution of any functionally independent statistic T is independent of the population (supposed normal) from which the sample has been drawn.

The members of this population A of samples differ among themselves in the values, μ and σ , of the means and standard deviations of the parent populations from which they are drawn. In designing a test involving the mean only we require a test or pivotal quantity, involving μ only, the distribution of which is wholly independent of the parameters. This is supplied by

$$t = (\bar{x}-\mu)\sqrt{N}/s$$

The distribution of this quantity depends on N, for the number of degrees of freedom is $n = N-1$. This is a matter of no consequence, for the function of t and N,

$$P(t) = \frac{\left(\frac{n-1}{2}\right)!}{\left(\frac{n-2}{2}\right)!} \; \frac{1}{\sqrt{\pi n}} \int_t^\infty \left(1+\frac{u^2}{n}\right)^{-(1+n)/2} du, \qquad \ldots (2)$$

might equally well have been chosen as the pivotal quantity, and its distribution $0 \leqslant P \leqslant 1$, $df = dP$, is independent of N also. The essential step which we now take is to assert that, whatever may be the character of our sample, we shall derive the probability that μ shall fall in any range from the known probability that P and t shall fall in the ranges corresponding for these quantities. It is in this sense that we interpret the somewhat paradoxical statement, with which the problem at first confronts us, that a sample having *known* characteristics is a *random* sample of an unknown population. The properties of variable statistics derived from observations, defined as random variables with distribution functions given in terms of known parameters, are used to establish the connections, (1) and (2), which once established serve to give meaning to the situation which we encounter in practice, in which the statistics are observable, but the parameters unknown.

It is instructive to compare the general form of the fiducial argument set out above with a special case of great simplicity, suitable for examining its logical cogency.

Let μ be the median of a distribution of which nothing is known save that its probability integral is continuous (e.g. it need not be differentiable). Let x_1 and x_2 be two obeservations of the variate ; then for any given value of μ it will be true that

(1) in one case out of 4 both x_1 and x_2 will exceed the median,

(2) in two cases out of 4, one value will exceed and the other be less than the median,

(3) in one case out of 4, both will be less than the median.

If a stands for the number of observations less than the median, then a will be a pivotal quantity involving both the unknown parameter and the observations, and having a sampling distribution independent of the parameter, i.e. a takes the values 0, 1, and 2 with probabilities $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ respectively.

Recognising this property we may argue from two given observations, now regarded as fixed parameters that the probability is $\frac{1}{4}$ that μ is less than both x_1 and x_2, that the probability is $\frac{1}{2}$ that μ lies between x_1 and x_2, and that the probability is $\frac{1}{4}$ that μ exceeds both x_1 and x_2. The argument thus leads to a frequency distribution of μ, now regarded as a random variate.*

The idea that probability statements about unknown parameters cannot be derived from *data* consisting of observations can only be upheld by those willing to reject this simple argument.

I do not think that any other interpretation of the known as a random variable of the unknown would be rational and consistent. It is, however, worth while to endeavour to see how, long after the tests had become established in statistical practice, a misunderstanding of their logic should have been possible.

Now, the derivation of Student's distribution involves the removal of a variable by integration, in a way which happens to be capable of such misinterpretation. The relation between a sample from a normal population and the population from which it is drawn is expressible in terms of two independent random variables,

---

* I have, of course, given elsewhere (Obituary "Student") the general fiducial distribution for this problem, namely

$$N\,!\; dp_1\, dp_2 \ldots dp_N, \qquad p_1 > p_2 > \ldots > p_N$$

where $p_1\, p_2 \ldots p_N$ are pivotal quantities defind as the fractions of the population sampled exceeded by each of N observations in descending order of magnitude.

$$(\bar{x}-\mu)\sqrt{\mathrm{N}}\,/\,\sigma = \xi, \qquad \ldots \quad (3)$$

where ξ is a random variable, such that

$$\mathrm{P}(\xi)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\xi} e^{-\frac{1}{2}x^2}\,dx \qquad \ldots \quad (4)$$

and

$$\tfrac{1}{2}\,ns^2/\sigma^2=u$$

where u is a random variable, with the Eulerian distribution

$$\mathrm{P}(u)=\frac{1}{\left(\frac{n-2}{2}\right)!}\int_{0}^{u} x^{(n-2)/2}\,e^{-x}\,dx.$$

Both random variables involve the unknown parameter σ, and to eliminate this we put

$$t=(\bar{x}-\mu)\sqrt{\mathrm{N}}/s=\xi\sqrt{n/2u}, \quad \text{so that} \quad \tfrac{1}{2}\xi^2=(t^2/n)u,$$

and a random variable independent of σ may be obtained by integrating, with respect to u the simultaneous frequency element

$$\frac{1}{\sqrt{2\pi}\left(\frac{n-2}{2}\right)!}\,e^{-u(1+t^2/n)}\,\sqrt{2u/n.}\;u^{(n-2)/2}\,du\,dt$$

leading to the familiar distribution of t. It so happens in this case, though it is of no advantage to the test of significance, that, with the unknown σ, the known quantity s is also eliminated. The misunderstanding to which I refer is that which represents the elimination of s as the purpose of the integration, as though we needed to trace out the frequency distribution of t in successive samples from a population with σ given. This is wholly illogical since s and not σ is in fact used in making the test.

Such a mistake is of only academic importance in the simple case of Student's test, but becomes a real hindrance to understanding the more complex tests based upon it. When, for example, we have samples from two different populations, their two unknown variances are both eliminated, giving the simultaneous distribution of two independent values of t. To test whether the means could be equal, if μ stand for the common mean, we have

$$\bar{x}_1-\mu=s_1t_1, \quad \bar{x}_2-\mu=s_2t_2, \quad \text{whence } \bar{x}_1-\bar{x}_2=s_1t_1-s_2t_2,$$

and, knowing the distribution of t_1 and t_2 we are in a position to calculate what value of $\bar{x}_1-\bar{x}_2$ will be exceeded with any chosen frequency. The view that the purpose of the integration was to eliminate s_1 and s_2 would naturally require that the frequency distribution of $\bar{x}_1-\bar{x}_2$ whould be determined for all sampling values of s_1 and s_2, i.e. as a normal distribution with variance $\sigma_1{}^2/\mathrm{N}_1+\sigma_2{}^2/\mathrm{N}_2$ so reintroducing the unknown σ_1 and σ_2. Reference to σ_1 and σ_2 puts the argument back to the stage for which Bayes method was originally developed.

The population of samples of which some standard fraction has a value of $\bar{x}_1-\bar{x}_2$ exceeding the level found for significance is not, however, a population of samples drawn in succession from some fixed population of unknown characteristics, but the population of samples (identical with the observed in the whole set of statistics $\mathrm{N}_1, \mathrm{N}_2, s_1, s_2$) determined by the fundamental random-variable relation by which sample and population are connected.

Paper received : 18 *April*, 1945.

THE INDIVIDUAL SAMPLING DISTRIBUTION OF THE MAXIMUM, THE MINIMUM AND ANY INTERMEDIATE OF THE p-STATISTICS ON THE NULL-HYPOTHESIS

By SAMARENDRA NATH ROY
Statistical Laboratory, Calcutta.

INTRODUCTION

Given two random samples S$'$ and S$''$ of sizes n' and n'', and variance and covariance matrices $\|a'_{ij}\|$ and $\|a''_{ij}\|$ drawn from two p-variate normal populations Σ' and Σ'' with variance and covariance matrices $\|\alpha'_{ij}\|$ and $\|\alpha''_{ij}\|$, it was shown earlier by the author (1939) and also by others (1939) that the null hypothesis (for the populations) $\|\alpha'_{ij}\|=\|\alpha''_{ij}\|$ could be appropriately tested by a set of p-statistics given by the p roots $k^2_1, k^2_2, \ldots, k^2_p$ (all positive in this particular case) of the determinantal equation in $k^2 \rightarrow |a'_{ij} - k^2 a''_{ij}| = 0$. The joint sampling distribution of these p-statistics on the null hypothesis $\|\alpha'_{ij}\|=\|\alpha''_{ij}\|$ was at the same time obtained by the author (1939) and by others (1939). The joint distribution of these k_i's$(i=1, 2, \ldots, p)$ on the non-null hypothesis (that is, when $\|\alpha'_{ij}\| \neq \|\alpha''_{ij}\|$) was also given by the author a little later (1942) and it was found that this involved as population parameters p quantities $\kappa^2_1, \kappa^2_2, \ldots, \kappa^2_p$ which are all unity when and only when $\|\alpha'_{ij}\| = \|\alpha''_{ij}\|$ and which come out as the p roots (all positive in this case) of the determinantal equation in $\kappa^2 \rightarrow |\alpha'_{ij} - \kappa^2 \alpha''_{ij}| = 0$. Closely associated with this is the second problem of multivariate analysis of variance which can be stated as follows. Given l samples $S_1, S_2, \ldots, S_l$ of sizes $n_1, n_2, \ldots, n_l$ and mean values $\bar{x}_i(r)$ and variance and covariance matrices $\|a_{ij}(r)\|$ (where i, j refer to the different variates, and r to the sample, and $i, j, = 1, 2, \ldots, p$ and $r = 1, 2, \ldots, l$) supposed to have been drawn from l p-variate normal populations $\Sigma_1, \Sigma_2, \ldots, \Sigma_l$ with mean values $m_i(r)$ and variance and covariance matrix $\|\beta''_{ij}\|$ (supposed to be the same for all populations), to test the null hypothesis $m_i(1)=m_i(2) = \ldots = m_i(l)$, $(i=1, 2, \ldots, p)$. This is, of course, in terms of one-way classification. The same technique could by a slight twist be used for analysis of variance in terms of multiway and other types of classifications. It was shown by the author (1939) that the null hypothesis in the case just mentioned could be appropriately tested (when $l > p$) by a set of p-statistics $t^2_1, t^2_2, \ldots, t^2_p$ which come out as the p roots (all positive in this case) of the determinantal equation in $t^2 \rightarrow |b'_{ij} - t^2 b''_{ij}| = 0$, where

$$\left.\begin{aligned} b'_{ij} &= \sum_{r=1}^{l} n_r\{\bar{x}_i(r)-\bar{x}_i\}\ \{\bar{x}_j(r)-\bar{x}_j\}/(l-1)\ ; \quad b''_{ij}= \sum_{=r1}^{l} (n-1)a_{ij}(r)/ \sum_{r=1}^{l} (n_r-1) \\ \bar{x}_i &= \sum_{r=1}^{l} n_r\, \bar{x}_i(r) / \sum_{r=1}^{l} n_r, \quad (i, j=1, 2, \ldots, p) \end{aligned}\right\} \quad \ldots \quad (1.1)$$

The joint distribution of these p-statistics $t_1, t_2, \ldots, t_p$ on the null-hypothesis $m_i(1) = m_i(2) = \ldots =, m_i(l)$ $(i=1, 2, \ldots, p)$ comes out to be of the same form as that of $k_1, k_2, \ldots, k_p$ of the first problem and was indicated as such by the author (1939) and by others (1939). The distribution on the non-null hypothesis $m_i(1) \neq m_i(2) \neq \ldots \neq m_i(l)$ $(i=1, 2, \ldots, p)$ was given by the author (1942) a little later and involved as parameter a certain function of p quantities $\tau_1, \tau_2, \ldots, \tau_p$ where $\tau^2_1, \tau^2_2, \ldots, \tau^2_p$ are the p roots (all positive in this case) of the determinantal equation in $\tau^2 \rightarrow |\beta'_{ij} - \tau^2 \beta''_{ij}| = 0$ where β''_{ij} is already defined and

$$\beta'_{ij} = \sum_{r=1}^{l} n_r \{m_i(r) - \bar{m}_i\} \{m_j(r) - \bar{m}_j\} / (l-1) \; ; \; \bar{m}_i = \sum_{r=1}^{l} n_r \, m_i(r) / \sum_{r=1}^{l} n_r \qquad \ldots \quad (1.2)$$

It was noted that τ_i's $(i = 1, 2, \ldots, p)$ are zero when and only when $\| \beta'_{ij} \| = 0$ which happens when and only when $m_i(1) = m_i(2) = \ldots m_i(l)$, $(i = 1, 2, \ldots, p)$. It is also worth noting here that while on the respective null hypotheses the joint distribution of k_i's of the first problem has the same form as that of t_i's of the second problem, the corresponding distributions on the respective non-null hypotheses are entirely different. This feature of the problem is well-known in the univariate case and is found to be carried over into the multivariate in a manner one would be naturally led to expect. The case for $l \leqslant p$ has also been since investigated and the solution will be out shortly.

In the first case on the null hypothesis the joint sampling distribution of the p statistics $k^2_1, k^2_2, \ldots, k^2_p$ (supposed to be arranged in an ascending order of magnitude) comes out as

$$\text{Const.} \prod_{=1}^{p} k_i^{n'-p-1} \, dk_i / (1 + \lambda k^2_i)^{\frac{n'+n''-2}{2}}$$

$$\times \begin{vmatrix} (k^2_p)^{p-1} & (k^2_{p-1})^{p-1} & \ldots & (k^2_2)^{p-1} & (k^2_1)^{p-1} \\ (k^2_p)^{p-2} & (k^2_{p-1})^{p-2} & \ldots & (k^2_2)^{p-2} & (k^2_1)^{p-2} \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ k^2_p & k^2_{p-1} & \ldots & k^2_2 & k^2_1 \\ 1 & 1 & \ldots & 1 & 1 \end{vmatrix} \qquad \ldots \quad (1.3)$$

where

$$\lambda = (n'-1)/(n''-1) \qquad \ldots \quad (1.4)$$

while in the second case on the null hypothesis the distribution of $t_1, t_2, \ldots, t_p$ (supposed to be arranged in an ascending order of magnitude) comes out as

$$\text{Const} \prod_{i=1}^{p} t_i^{l-p-1} \, dt_i / (1 + \mu \, t^2_i)^{\frac{N-l}{2}}$$

$$\times \begin{vmatrix} (t^2_p)^{p-1} & (t^2_{p-1})^{p-1} & \ldots & (t^2_2)^{p-1} & (t^2_1)^{p-1} \\ (t^2_p)^{p-2} & (t^2_{p-1})^{p-2} & \ldots & (t^2_2)^{p-2} & (t^2_1)^{p-2} \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ t^2_p & t^2_{p-1} & \ldots & t^2_2 & t^2_1 \\ 1 & 1 & \ldots & 1 & 1 \end{vmatrix} \qquad \ldots \quad (1.5)$$

where

$$\mu = (l-1)/(N-l) \text{ and } N = \sum_{r=1}^{l} n_r \qquad \ldots \quad (1.6)$$

It will be seen that (1.3) and (1.5) are, as noted earlier, exactly of the same form.

The object of the present paper is to find out the sampling distribution of (*i*) the maximum statistic k_p or t_p of (1.3) or (1.5), (*ii*) the minimum statistic k_1 or t_1 of (1.3) or (1.5), and (*iii*) in general any intermediate statistic k_j or t_j $(j = 2, 3, \ldots, p-1)$. It is evident that one line of investigation would serve for both k's and t's. The next paper will be concerned with the sampling distribution of (*i*) k_1, k_p or $k_j (j = 2, 3, \ldots, p-1)$ on the non-null hypothesis $\|\alpha'_{ij}\| \neq \|\alpha''_{ij}\|$ and (*ii*) t_1, t_p and $t_j (j = 2, 3, \ldots, p-1)$ on the non-null hypothesis $\|\beta'_{ij}\| \neq 0$.

2. NOTATION AND MATHEMATICAL PRELIMINARIES

If in (1.3) we put $\lambda k^2_i = x'_i$ $(i = 1, 2, \ldots, p)$, then (1.3) reduces to

$$\text{Const.} \prod_{i=1}^{p} x'^{\frac{n'-p-2}{2}}_i \, dx'_i/(1+x'_i)^{\frac{n'+n''-2}{2}}$$

$$\times \begin{vmatrix} (x'_p)^{p-1} & (x'_{p-1})^{p-1} & \cdots & (x'_2)^{p-1} & (x'_1)^{p-1} \\ (x'_p)^{p-2} & (x'_{p-1})^{p-2} & \cdots & (x'_2)^{p-2} & (x'_1)^{p-2} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ x'_p & x'_{p-1} & \cdots & x'_2 & x'_1 \\ 1 & 1 & \cdots & 1 & 1 \end{vmatrix} \quad \ldots (2{\cdot}1)$$

Similarly if we put $\mu t^2_i = x''_i (i = 1, 2, \ldots p)$ (1.5) goes over into

$$\text{Const.} \prod_{i=1}^{p} x''^{\frac{l-p-2}{2}} \, dx''_i/(1+x''_i)^{\frac{N-1}{2}}$$

$$\times \begin{vmatrix} (x''_p)^{p-1} & (x''_{p-1})^{p-1} & \cdots & (x''_2)^{p-1} & (x''_1)^{p-1} \\ (x''_p)^{p-2} & (x''_{p-1})^{p-2} & \cdots & (x''_2)^{p-2} & (x''_1)^{p-2} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ x''_p & x''_{p-1} & \cdots & x''_2 & x''_1 \\ 1 & 1 & \cdots & 1 & 1 \end{vmatrix} \quad \ldots (2.2)$$

Now (2.1) and (2.2) being of the same form we can conveniently put x_i for either x'_i or x''_i, m for either $(n'-p-2)/2$ or $(l-p-2)/2$, and n for either $(n'+n''-2)/2$ or $(N-1)/2$. It is to be noticed that $(n'-p-2)/2 < (n'+n''-2)/2$ and $(l-p-2)/2 < (N-1)/2$, which means that m in any case is less than n. We can now write (2.1) or (2.2) in the form

$$\text{Const.} \prod_{i=1}^{p} x^{m}_i \, dx_i/(1+x_i)^{n}$$

$$\times \begin{vmatrix} x_p^{p-1} & x_{p-1}^{p-1} & \cdots & x_2^{p-1} & x_1^{p-1} \\ x_p^{p-2} & x_{p-1}^{p-2} & \cdots & x_2^{p-2} & x_1^{p-2} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ x_p & x_{p-1} & \cdots & x_2 & x_1 \\ 1 & 1 & \cdots & 1 & 1 \end{vmatrix} \quad \ldots (2.3)$$

It was shown in earlier papers that, unarranged, each of the k'_is or t'_is varies from 0 to ∞ and arranged in an ascending order we have

$$0 \leqslant k_1 \leqslant k_2 \leqslant \ldots \leqslant k_{p-1} \leqslant k_p \leqslant \infty, \text{ and } 0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_{p-1} \leqslant t_p \leqslant \infty \quad \ldots (2.4)$$

It is to be remembered here that (1.3) and (1.5) or (2.1) and (2.2) give the relevant distributions only when the statistics have been arranged in an ascending order. Otherwise, as pointed out in earlier papers, instead of the determinants occurring here we would have had absolute values of these determinants coming in.

Now from (2.4) it easily follows that

$$0 \leqslant x_1 \leqslant x_2 \leqslant \ldots \leqslant x_{p-2} \leqslant x_{p-1} \leqslant x_p \leqslant \infty \qquad \ldots \quad (2.41)$$

The distribution (2.3) can be conveniently rewritten in the form

$$\text{Const.} \prod_{i=1}^{p} dx_i$$

$$\times \begin{vmatrix} \dfrac{x_p{}^{m+p-1}}{(1+x_p)^n} & \dfrac{x_{p-1}{}^{m+p-1}}{(1+x_{p-1})^n} & \cdots & \dfrac{x_2{}^{m+p-1}}{(1+x_2)^n} & \dfrac{x_1{}^{m+p-1}}{(1+x_1)^n} \\ \dfrac{x_p{}^{m+p-2}}{(1+x_p)^n} & \dfrac{x_{p-1}{}^{m+p-2}}{(1+x_{p-1})^n} & \cdots & \dfrac{x_2{}^{m+p-2}}{(1+x_2)^n} & \dfrac{x_1{}^{m+p-2}}{(1+x_1)^n} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \dfrac{x_p{}^{m+1}}{(1+x_p)^n} & \dfrac{x_{p-1}{}^{m+1}}{(1+x_{p-1})^n} & \cdots & \dfrac{x_2{}^{m+1}}{(1+x_2)^n} & \dfrac{x_1{}^{m+1}}{(1+x_1)^n} \\ \dfrac{x_p{}^{m}}{(1+x_p)^n} & \dfrac{x^{m}{}_{p-1}}{(1+x_{p-1})^n} & \cdots & \dfrac{x_2{}^{m}}{(1+x_2)^n} & \dfrac{x_1{}^{m}}{(1+x_1)^n} \end{vmatrix} \qquad \ldots \quad (2.42)$$

or in a more compressed notation

$$\text{Const. } \mathrm{F}\{m_p, n\,;\, x_p : m_{p-1}, n\,;\, x_{p-1} : \ldots .\, m_2, n\,;\, x_2 : m_1, n\,;\, x_1\} \times \prod_{i=1}^{p} dx_i \qquad \ldots \quad (2.43)$$

where the function F{ } stands for the determinant in (2.42) and where $m_i = m + i - 1$ $(i = p, p-1 \ldots 1)$

The incomplete probability integrals of the distributions of $(i) x_p$, $(ii) x_1$, and (iii) x_s for $(s=p-1, \ldots, 2)$ are given respectively by the p-fold integrals

$$(i) \int_0^{x} dx_p \int_0^{x_p} dx_{p-1} \int_0^{x_{p-1}} dx_{p-2} \ldots \int_0^{x_2} dx_1 \mathrm{F}\{m_p\, n\,;\, x_p : m_{p-1}, n;\, x_{p-1} : \ldots m_1\, n;\, x_1\} \qquad \ldots \quad (2.5)$$

$$(ii) \int_{x_{p-1}}^{\infty} dx_p \int_{x_{p-2}}^{\infty} dx_{p-1} \ldots \int_{x_1}^{\infty} dx_2 \int_{x}^{\infty} dx_1 \mathrm{F}\{m_p, n;\, x_p : m_{p-1}, n;\, x_{p-1} : \ldots m_1, n;\, x_1\} \qquad \ldots \quad (2.51)$$

$$(iii) \int_{x_{p-1}}^{\infty} dx_p \int_{x_{p-2}}^{\infty} dx_{p-1} \ldots \int_{x_j}^{\infty} dx_{j+1} \int_{x}^{\infty} dx_j \int_0^{x_j} dx_{j-1} \ldots \int_0^{x_2} dx_1 \mathrm{F}\{m_p\, n;\, x_p \ldots m_1, n;\, x_1\} \qquad (2.52)$$

In (i) the order of integration is from $x_1 \to x_2 \to x_3 \to \ldots x_p$; in (ii) the order is from $x_p \to x_{p-1} \to x_{p-2} \to \ldots \to x_1$; in (iii) it is from $x_p \to x_{p-1} \to \ldots \to x_j$ on one side and from $x_1 \to x_2 \to \ldots \to x_j$ on the other side. It will be also seen that (i) gives the incomplete probability integral of x_p from 0 to x whence we could, if we liked, easily obtain the incomplete integral from x to ∞, (ii) gives the incomplete probability integral of x_1 from x to ∞ whence that for the range 0 to x could be easily obtained, while (iii) gives the incomplete integral for x_j from x to ∞ whence again that from 0 to x is immediately derivable.

In the function F{ } defined by (2.43) we have the restriction (*i*) $\infty \geqslant x_p \geqslant x_{p-1} \geqslant x_{p-2} \geqslant \ldots \geqslant x_2 \geqslant x_1 \geqslant 0$, (*ii*) $m_s = m+s-1 (s=p, p-1, \ldots, 1)$ and (*iii*) all the n's are equal and greater than m_p. It will be necessary and hence desirable to relax (*ii*) and (*iii*) and consider $m_p, n_p\,;\ m_{p-1}, n_{p-1}\,;\ \ldots m_1, n_1$ subject, however, to the provisos that $(\alpha)\ m_p > m_{p-1} > \ldots > m_1$, $(\beta) n_p > n_{p-1} > \ldots > n_1$ and $(\gamma) n_1 > m_p$. The more general type of function F{ } will then be

$$F\{m_p, n_p : x_p : m_{p-1}, n_{p-1} ; x_{p-1} : \ldots m_2, n_2 ; x_2 : m_1, n_1 : x_1\} \qquad \ldots (2.6)$$

where the x_s's are subject to (*i*) and m_s's and n_s's are subject to (α), (β) and (r), and $s=p, p-1, \ldots, 2, 1$. For such a function we shall have, of course, incomplete probability integrals corresponding to (2.5), (2.51) and (2.52). Such integrals would not directly arise out of the statistical situations considered here, but it will be seen later on that in course of reduction of the main integrals, such integrals would occur as mathematical auxiliaries not open to direct statistical interpretation.

We shall denote (2.5), (2.51) and (2.52) respectively by

$$\left.\begin{array}{l} F\{x ; m_p, n ; m_{p-1}\, n ; \ldots m_1, n\}, \\ F\{m_p, n ; m_{p-1}, n ; \ldots m_1, n ; x\} \text{ and} \\ F\{m_p, n : \ldots m_j, n : x ; m_{j-1}, n ; \ldots m_1, n\} \end{array}\right\} \qquad \ldots (2.7)$$

Or, since all the n's are equal, by

$$\left.\begin{array}{l} F\{x ; m_p, m_{p-1}, \ldots m_1 ; n\}, \\ F\{m_p, m_{p-1}, \ldots m_1 ; n ; x\} \text{ and} \\ F\{m_p, m_{p-1}, \ldots m_j ; x ; m_{j-1}, \ldots m_1 ; n\} \end{array}\right\} \qquad \ldots (2.71)$$

while the corresponding incomplete integrals arising out of (2.6) can be, of course, conveniently denoted by

$$F\{x ; m_p, n_p ; m_{p-1}, n_{p-1} ; \ldots m_1, n_1,\},\ F\{m_p, n_p ; m_{p-1}, n_{p-1} ; \ldots m_1, n_1 ; x\}$$

and

$$F\{m_p, n_p ; \ldots m_j, n_j ; x ; m_{j-1}, n_{j-1} ; \ldots m_1, n_1\} \qquad \ldots (2.72)$$

It will be seen from (2.42), (2.43), (2.5), (2.52) and (2.6) that the evaluation of the functions defined by (2.7)—(2.72) involves a consideration of the following types of integrals

$$\int_0^{x} \frac{x_p^{m'_p}\, dx_p}{(1+x_p)^{n'_p}} \int_0^{x_p} \frac{x_{p-1}^{m'_{p-1}}\, dx_{p-1}}{(1+x_{p-1})^{n'_{p-1}}} \cdots \int_0^{x_2} \frac{x_1^{m'_1} dx_p}{(1+x_1)^{n'_1}}\ ;\ \int_{x_{p-1}}^{\infty} \frac{x_p^{m'_p}\, dx_p}{(1+x_p)^{n'_p}} \cdots \int_{x}^{\infty} \frac{x_1^{m'}\, dx_1}{(1+x_1)^{n'_1}}$$

and

$$\int_{x_{p-1}}^{\infty} \frac{x_p^{m'_p}\, dx_p}{(1+x_p)^{n'_p}} \cdots \int_{x_j}^{\infty} \frac{x_{j+1}^{m'_{j+1}}\, dx_{j+1}}{(1+x_{j+1})^{n'_{j+1}}} \int_{x}^{\infty} \frac{x_j^{m'_j}\, dx_j}{(1+x_j)^{n'_j}} \int_0^{x_j} \frac{x_{j-1}^{m'_{j-1}}\, dx_{j-1}}{(1+x_{j-1})^{n'_{j-1}}} \cdots \int_0^{x_2} \frac{x_1^{m'_1}\, dx_1}{(1+x_1)^{n'_1}} \qquad \ldots (2.73)$$

where $(m'_p, m'_{p-1}, \ldots m'_1)$ is any permutation of $m_p, m_{p-1}, \ldots m_1$ and $(n'_p, \ldots n'_1)$ is the corresponding permutation of $(n_p, n_{p-1}, \ldots n_1)$.

The integrals given by (2.73) we can conveniently denote by

$$F(x ; m'_p, n'_p ; m'_{p-1}, n'_{p-1} ; \ldots m'_1, n'_1) ;\ F(m'_p, n'_p ; m'_{p-1}, n'_{p-1} ; m'_1, n'_1 ; x)$$

and

$$F(m'_p, n'_p ; \ldots ; m'_j, n'_j ; x ; m'_{j-1}, n'_{j-1} ; \ldots ; m'_1, n'_1) \qquad \ldots (2.74)$$

The functions F{ } defined earlier and F( ) defined now should be carefully distinguished. For purposes of reduction and demonstration another alternative notation to (2.7), (2.71) and (2.72) appears to be more helpful namely

$$\mathrm{F}\left\{ \mathrm{x}\, ; \begin{pmatrix} m_p,\, n_p & m_{p-1},\, n_{p-1} & \cdots & m_1,\, n_1 \\ m_p,\, n_p & m_{p-1},\, n_{p-1} & \cdots & m_1,\, n_1 \\ \cdot & \cdot & \cdots & \cdot \\ m_p,\, n_p & m_{p-1},\, n_{p-1} & \cdots & m_1,\, n_1 \end{pmatrix} \right\},\quad \mathrm{F}\left\{ \begin{pmatrix} m_p,\, n_p & m_{p-1},\, n_{p-1} & \cdots & m_1,\, n_1 \\ m_p,\, n_p & m_{p-1},\, n_{p-1} & \cdots & m_1,\, n_1 \\ \cdot & \cdot & \cdots & \cdot \\ m_p,\, n_p & m_{p-1},\, n_{p-1} & \cdots & m_1,\, n_1 \end{pmatrix} ; \mathrm{x} \right\}$$

$$\text{and } \mathrm{F}\left\{ \begin{matrix} m_p,\, n_p & m_{p-1},\, n_{p-1} & \cdots & m_1,\, n_1 \\ \cdot & \cdot & & \cdot \\ m_p,\, n_p & m_{p-1},\, n_{p-1} & \cdots & m_1,\, n_1 \\ \cdot & \cdot & \cdots & \cdot \\ \mathrm{x};\, (m_p,\, n_p & m_{p-1},\, n_{p-1} & \cdots & m_1,\, n_1) \end{matrix} \right\} \qquad \ldots \quad (2.75)$$

The notation can be illustrated for the case of three variates for which

$$\mathrm{F}\left\{ \mathrm{x}\, ; \begin{pmatrix} m_3,\, n_3 & m_2,\, n_2 & m_1,\, n_1 \\ m_3,\, n_3 & m_2,\, n_2 & m_1,\, n_1 \\ m_3,\, n_3 & m_2,\, n_2 & m_1,\, n_1 \end{pmatrix} \right\}$$

$$\equiv + \int_0^{\mathrm{x}} \frac{x_3^{m_3}\, dx_3}{(1+x_3)^{n_3}} \left(\int_0^{x_3} \frac{x_2^{m_2}\, dx_2}{(1+x_2)^{n_2}} \int_0^{x_2} \frac{x_1^{m_1}\, dx_1}{(1+x_1)^{n_1}} - \int_0^{x_3} \frac{x_2^{m_1}\, dx_2}{(1+x_2)^{n_1}} \int_0^{x_2} \frac{x_1^{m_2}\, dx_1}{(1+x_1)^{n_2}} \right)$$

$$- \int_0^{\mathrm{x}} \frac{x_3^{m_2}\, dx_3}{(1+x_3)^{n_2}} \left(\int_0^{x_3} \frac{x_2^{m_3}\, dx_2}{(1+x_2)^{n_3}} \int_0^{x_2} \frac{x_1^{m_1}\, dx_1}{(1+x_1)^{n_1}} - \int_0^{x_3} \frac{x_2^{m_1}\, dx_2}{(1+x_2)^{n_1}} \int_0^{x_2} \frac{x_1^{m_3}\, dx_1}{(1+x_1)^{n_3}} \right)$$

$$+ \int_0^{\mathrm{x}} \frac{x_3^{m_1}\, dx_3}{(1+x_3)^{n_1}} \left(\int_0^{x_3} \frac{x_2^{m_3}\, dx_2}{(1+x_2)^{n_3}} \int_0^{x_2} \frac{x_1^{m_2}\, dx_1}{(1+x_1)^{n_2}} - \int_0^{x_3} \frac{x_2^{m_2}\, dx_2}{(1+x_2)^{n_2}} \int_0^{x_2} \frac{x_1^{m_3}\, dx_1}{(1+x_1)^{n_3}} \right)$$

$$\equiv \mathrm{F}(\mathrm{x};\, m_3,\, n_3;\, m_2,\, n_2;\, m_1,\, n_1;) - \mathrm{F}(\mathrm{x};\, m_3,\, n_3;\, m_1,\, n_1;\, m_2,\, n_2) - \mathrm{F}(\mathrm{x};\, m_2,n_2;\, m_3,n_3;\, m_1,n_1)$$

$$+ \mathrm{F}(\mathrm{x};\, m_2,\, n_2;\, m_1,\, n_1;\, m_3,\, n_3;) + \mathrm{F}(\mathrm{x};\, m_1,\, n_1;\, m_3,\, n_3;\, m_2,\, n_2;)$$

$$- \mathrm{F}(\mathrm{x};\, m_1,\, n_1;\, m_2,\, n_2;\, m_3,\, n_3) \qquad \ldots \quad (2.76)$$

Similarly,

$$\mathrm{F}\left\{ \begin{pmatrix} m_3,\, n_3 & m_2,\, n_2 & m_1,\, n_1 \\ m_3,\, n_3 & m_2,\, n_2 & m_1,\, n_1 \\ m_3,\, n_3 & m_2,\, n_2 & m_1,\, n_1 \end{pmatrix} ;\ \mathrm{x} \right\}$$

$$= + \int_{x_2}^{\infty} \frac{x_3^{m_3}\, dx_3}{(1+x_3)^{n_3}} \left(\int_{x_1}^{\infty} \frac{x_2^{m_2}\, dx_2}{(1+x_2)^{n_2}} \int_{\mathrm{x}}^{\infty} \frac{x_1^{m_1}\, dx_1}{(1+x_1)^{n_1}} - \int_{x_1}^{\infty} \frac{x_2^{m_1}\, dx_2}{(1+x_2)^{n_1}} \int_{\mathrm{x}}^{\infty} \frac{x_1^{m_2}\, dx_1}{(1+x_1)^{n_2}} \right)$$

$$- \int_{x_2}^{\infty} \frac{x_3^{m_2}\, dx_3}{(1+x_3)^{n_2}} \left(\int_{x_1}^{\infty} \frac{x_2^{m_3}\, dx_2}{(1+x_2)^{n_3}} \int_{\mathrm{x}}^{\infty} \frac{x_1^{m_1}\, dx_1}{(1+x_1)^{n_1}} - \int_{x_1}^{\infty} \frac{x_2^{m_1}\, dx_2}{(1+x_2)^{n_1}} \int_{\mathrm{x}}^{\infty} \frac{x_1^{m_3}\, dx_1}{(1+x_1)^{n_3}} \right)$$

$$+ \int_{x_2}^{\infty} \frac{x_3^{m_1}\, dx_3}{(1+x_3)^{n_1}} \left(\int_{x_1}^{\infty} \frac{x_2^{m_3}\, dx_2}{(1+x_2)^{n_3}} \int_{\mathrm{x}}^{\infty} \frac{x_1^{m_2}\, dx_1}{(1+x_1)^{n_2}} - \int_{x_1}^{\infty} \frac{x_2^{m_2}\, dx_2}{(1+x_2)^{n_2}} \int_{\mathrm{x}}^{\infty} \frac{x_1^{m_3}\, dx_1}{(1+x_1)^{n_3}} \right)$$

$$= \mathrm{F}(m_3,n_3;\, m_2,n_2;\, m_1,n_1;\, \mathrm{x}) - \mathrm{F}(m_3,n_3;\, m_1,n_1;\, m_2,n_2;\, \mathrm{x}) - \mathrm{F}(m_2,n_2;\, m_3,n_3;\, m_1,n_1;\, \mathrm{x})$$

$$+ \mathrm{F}(m_2,n_2;\, m_1,n_1;\, m_3 n_3;\, \mathrm{x}) + \mathrm{F}(m_1,n_1;\, m_3,n_3;\, m_2,n_2;\, \mathrm{x}) - \mathrm{F}(m_1,n_1;\, m_2,\, n_2;\, m_3,n_3;\, \mathrm{x}) \quad \ldots \ (2.77)$$

and

$$\mathrm{F}\left\{\begin{array}{l} \qquad m_3, n_3 \quad m_2, n_2 \quad m_1, n_1 \\ \qquad m_3, n_3 \quad m_2, n_2 \quad m_1, n_1 \\ \mathrm{x}\,;\ (m_3, n_3 \quad m_2, n_2 \quad m_1, n_1) \end{array}\right\}$$

$$\equiv +\int_{x_2}^{\infty}\frac{x_3^{m_3}\,dx_3}{(1+x_3)^{n_3}}\left(\int_{\mathrm{x}}^{\infty}\frac{x_2^{m_2}\,dx_2}{(1+x_2)^{n_2}}\int_0^{x_2}\frac{x_1^{m_1}\,dx_1}{(1+x_1)^{n_1}}-\int_{\mathrm{x}}^{\infty}\frac{x_2^{m_1}\,dx_2}{(1+x_2)^{n_1}}\int_0^{x_2}\frac{x_1^{m_2}\,dx_1}{(1+x_1)^{n_2}}\right)$$

$$-\int_{x_2}^{\infty}\frac{x_3^{m_2}\,dx_3}{(1+x_3)^{n_2}}\left(\int_{\mathrm{x}}^{\infty}\frac{x_2^{m_3}\,dx_2}{(1+x_2)^{n_3}}\int_0^{x_2}\frac{x_1^{m_1}\,dx_1}{(1+x_1)^{n_1}}-\int_{\mathrm{x}}^{\infty}\frac{x_2^{m_1}\,dx_2}{(1+x_2)^{n_1}}\int_0^{x_2}\frac{x_1^{m_3}\,dx_1}{(1+x_1)^{n_3}}\right)$$

$$+\int_{x_2}^{\infty}\frac{x_3^{m_1}\,dx_3}{(1+x_3)^{n_1}}\left(\int_{\mathrm{x}}^{\infty}\frac{x_2^{m_3}\,dx_2}{(1+x_2)^{n_3}}\int_0^{x_2}\frac{x_1^{m_2}\,dx_1}{(1+x_1)^{n_2}}-\int_{\mathrm{x}}^{\infty}\frac{x_2^{m_2}\,dx_2}{(1+x_2)^{n_2}}\int_0^{x_2}\frac{x_1^{m_3}\,dx_1}{(1+x_1)^{n_3}}\right)$$

$$=\mathrm{F}(m_3,n_3;\ m_2,n_2;\ \mathrm{x};\ m_1,n_1)-\mathrm{F}(m_3,n_3;\ m_1,n_1;\ \mathrm{x};\ m_2,n_2)-\mathrm{F}(m_2,n_2;\ m_3,n_3;\ \mathrm{x};\ m_1,n_1)$$
$$+\mathrm{F}(m_2,n_2;\ m_1,n_1;\ \mathrm{x};\ m_3,n_3)+\mathrm{F}(m_1,n_1;\ m_3,n_3;\ \mathrm{x};\ m_2,n_2)-\mathrm{F}(m_1,n_1;\ m_2,n_2;\ \mathrm{x};\ m_3,n_3) \quad .. \ (2.78)$$

A detailed expansion for the case $p = 3$ makes evident the meaning of the notation for any general p. Going back to (2.42) we notice that in an actual p-variate problem $n_p = n_{p-1} = \ldots\ldots = n_1 = n$(suppose) and $m_s = m+s-1$ $(s=p,\ p-1, \ldots 1)$ although as we shall see later on, even there, for purposes of reduction the auxiliary functions F{ } have to be considered in which $n_p \neq n_{p-1} \neq \ldots n_1$ and $m_s \neq m+s-1$. If, however, $n_p = n_{p-1} = \ldots = n_1 = n$, then we can also more appropriately replace (2.75) by

$$\mathrm{F}\left\{\mathrm{x}\,;\begin{pmatrix} m_p & m_{p-1} & \cdot\cdot & m_1 \\ \cdot & \cdot & \cdot\cdot & \cdot \\ m_p & m_{p-1} & \cdot\cdot & m_1 \\ m_p & m_{p-1} & \cdot\cdot & m_1 \end{pmatrix}; n\right\};\ \mathrm{F}\left\{\begin{pmatrix} m_p & m_{p-1} & \cdot\cdot & m_1 \\ \cdot & \cdot & \cdot\cdot & \cdot \\ m_p & m_{p-1} & \cdot\cdot & m_1 \\ m_p & m_{p-1} & \cdot\cdot & m_1 \end{pmatrix}; n\,;\mathrm{x}\right\}$$

and

$$\mathrm{F}\left\{\begin{array}{l} \qquad m_p \quad m_{p-1} \quad \cdot\cdot \ m_1 \\ \qquad \ \cdot \qquad \cdot \qquad \cdot\cdot \ \ \cdot \\ \mathrm{x};\ (m_p \quad m_{p-1} \quad \cdot\cdot \ m_1)\ ;\ n \\ \qquad \ \cdot \qquad \cdot \qquad \cdot\cdot \ \ \cdot \\ \qquad m_p \quad m_{p-1} \quad \cdot\cdot \ m_1 \end{array}\right\} \quad .. \ (2.79)$$

The following is a group of fundamental theorems needed for reduction of the present problem

$$\int_0^{\mathrm{x}}\frac{x^m dx}{(1+x)^n}\,\mathrm{Q}(x)=-\frac{1}{n-m-1}\left[\frac{x^m}{(1+x)^{n-1}}\,\mathrm{Q}(x)\right]_0^{\mathrm{x}}+\frac{1}{n-m-1}\int_0^{\mathrm{x}}\frac{x^m\,dx}{(1+x)^{n-1}}\frac{\partial \mathrm{Q}}{\partial x}$$

$$+\frac{m}{n-m-1}\int_0^{\mathrm{x}}\frac{x^{m-1}\,dx}{(1+x)^n}\mathrm{Q}(x)=-\frac{1}{n-m-1}\,\frac{\mathrm{x}^m}{(1+\mathrm{x})^{n-1}}\,\mathrm{Q(x)}+\frac{1}{n-m-1}\int_0^{\mathrm{x}}\frac{x^m\,dx}{(1+x)^{n-1}}\,\frac{\partial \mathrm{Q}}{\partial x}$$

$$+\frac{m}{n-m-1}\int_0^{\mathrm{x}}\frac{x^{m-1}\,dx}{(1+x)^n}\mathrm{Q}(x) \quad .. \ (2{\cdot}8)$$

Here $\mathrm{Q}(x)$ is any function of x which conforms to the usual conditions of differentiability and integrability from 0 to ∞ and the result (2.8) is obtained by integrating $1/(1+x)^{n-m}$ and differentiating $\{x^m/(1+x)^m\}\,\mathrm{Q}(x)$. m is, of course, less than n as in all situations to be considered here. We can express (2.8) in a more convenient notation.

$$\mathrm{F}(\mathrm{x};\ m,n;\ \mathrm{Q})=-(1/n-m-1)\ \mathrm{F}_0(m,n-1;\ \mathrm{x})\mathrm{Q}(\mathrm{x})+(1/n-m-1)\ \mathrm{F}(\mathrm{x};\ m,n-1;\ \mathrm{Q}')$$
$$+(m/n-m-1)\mathrm{F}(\mathrm{x};\ m-1,\ n;\ Q) \quad .. \ (2{\cdot}81)$$

$$\left.\begin{array}{l} \text{where } \mathrm{F}(\mathrm{x};\ m,n;\ \mathrm{Q})=\displaystyle\int_0^{\mathrm{x}}\frac{x^m\,dx}{(1+x)^n}\,\mathrm{Q}(x),\ \ \mathrm{F}_0(m,n-1\ ;\ \mathrm{x})=\frac{\mathrm{x}^m}{(1+\mathrm{x})^{n-1}} \\ \text{and} \qquad \mathrm{Q}'=\dfrac{\partial \mathrm{Q}}{\partial x} \end{array}\right\} \quad .. \ (2.82)$$

It follows from (2.81) by transposition that

$$\begin{aligned}F(x; m,n; Q) = {} & (1/m+1)\ F_0(m+1,n-1; x)Q(x) - (1/m+1)\ F(x; m+1,n-1; Q') \\ & +(m+1/n-m-2)\ F(x; m+1,n; Q) \qquad .. \quad (2.83\end{aligned}$$

Similarly

$$\int_x^\infty \frac{x^m\,dx}{(1+x)^n} Q(x) = \frac{1}{n-m-1}\ \frac{x^m}{(1+x)^{n-1}}\ Q(x) + \frac{1}{n-m-1}\int_x^\infty \frac{x^m dx}{(1+)^{n-1}}\ \frac{\partial Q}{\partial x}$$

$$+\frac{m}{n-m-1}\int_x^\infty \frac{x^{m-1}\,dx}{(1+x)^n} Q(x)$$

or
$$\begin{aligned}F(m,n;\ Q; x) = {} & (1/n-m-1)\ F_0(m,\ n-1; x)Q(x) + (1/n-m-1)\ F(m,\ n-1;\ Q';\ x) \\ & +(m/n-m-1)\ F(m-1,\ n;\ Q;\ x) \qquad .. \quad (2.84)\end{aligned}$$

and by transposition

$$\begin{aligned}F(m,n\ ;\ Q; x) = {} & -(1/m+1)\ F_0(m+1,n-1; x)Q(x) - (1/m+1)\ F(m+1,n-1\ ;\ Q';\ x) \\ & +(n-m-2/m+1)\ F(m+1,n\ ;\ Q\ ;\ x) \qquad .. \quad (2.85)\end{aligned}$$

This is one group of theorems mostly needed here. There is another group which we should not directly use here but which might be useful in other types of reduction

$$\int_0^x \frac{x^m\,dx}{(1+x)^n} Q(x) = -\frac{1}{n-1}\left[\frac{x^m}{(1+x)^{n-1}}\ Q(x)\right]_0^x + \frac{1}{n-1}\int_0^x \frac{x^m\,dx}{(1+x)^{n-1}}\ \frac{\partial Q}{\partial x} + \frac{m}{n-1}\int_0^x \frac{x^{m-1}\,dx}{(1+x)^{n-1}}\ Q(x)$$

or
$$\left.\begin{aligned}F(x; m,n; Q) = {} & -(1/n-1)F_0(m,n-1;\ x)Q(x) + (1/n-1)\ F(x;\ m,n-1; Q') \\ & +(n/m+1)\ F(x;\ m+1,n-1; Q) \\ \text{and the associated theorem} & \\ F(x; m,n; Q) = {} & (1/m-1)\ F_0(m+1,n\ ; x)Q(x) - (1/m+1)\ F(x\ ;\ m+1,n; Q') \\ & +(n/m-1)\ F(x; m+1,n-1;\ Q)\end{aligned}\right\}\ ..\quad (2.86)$$

(2.86) is evidently obtained by integrating $1./(1+x)^n$ and differentiating $x^m Q(x)$. We have similarly

$$\left.\begin{aligned}F(m,n; Q; x) = {} & (1/n-1)\ F_0(m,n-1\ ;\ x)Q(x) + (1/n-1)\ F(m,n-1\ ;\ Q'\ ;\ x) \\ & +(m/n-1)\ F(m-1,n-1\ ;\ Q\ ;\ x) \\ \text{and alternatively} & \\ F(m,n\ ;\ Q\ ;\ x) = {} & -(1/m+1)\ F_0(m+1,n\ ;\ x)\ Q(x) - (1/m+1)\ F(m+1,n;\ Q\ ;\ x) \\ & +(n/m+1)\ F(m+1,n+1\ ;\ Q\ ;\ x)\end{aligned}\right\}\ ..\quad (2.87)$$

The way in which the theorems (2.8—2.85) are proposed to be used for purposes of reduction here may be illustrated as follows for the case of two variates ($p = 2$). With $n_1 = n_2 = n$ we have

$$\begin{aligned}F(x; m_2,n\ ;\ m_1,n) \equiv {} & \int_0^x \frac{x_2^{m_2}\ dx_2}{(1+x_2)^n} \int_0^{x_2} \frac{x_1^{m_1}\ dx_1}{(1+x)^n} \\ & = F(x\ ;\ m_2,\ n\ ;\ Q),\ \text{where}\ Q(x_1) = F(x_1; m_1, n) \qquad .. \quad (2.88)\end{aligned}$$

Remembering that $Q' = x_2^{m_1}/(1+x_2)^{n_1}$ we see that the right side of (2.88) reduces by using (2.81) to

$$\begin{aligned}&-(1/n-m_2-1)\,F_0(m_2,n_2-1\,;\,x)\,F(x\,;\,m_1,n)+(1/n-m_2-1)\,F(x\,;\,m_1+m_2;\,2n-1)\\&+(m_2/n-m_2-1)\,F(x\,;\,m_2-1\,;\,n\,;\,m_1,n).\end{aligned} \quad .. \quad (2.89)$$

This is if we integrate out by parts over x_2, trying to reduce m_2 in successive stages.

On the other hand we might integrate out by parts over x_1 trying to decrease m_1 in successive stages. In this case

$$F(x\,;\,m_2,n\,;\,m_1,n)=\int_0^x \frac{x_2^{m_2}\,dx_2}{(1+x_2)^n}\int_0^{x_2}\frac{x_1^{m_1}\,dx_1}{(1+x_1)^n}$$

$$\begin{aligned}&=\int_0^x \frac{x_2^{m_2}\,dx_2}{(1+x_2)^n}\left[-(1/n-m_1-1)\,\frac{x_2^{m_1}}{(1+x_2)^{n-1}}+(m_1/n-m_1-1)\int_0^{x_2}\frac{x_1^{m_1-1}\,dx_1}{(1+x_1)^n}\right]\\&=-(1/n-m_1-1)\,F(x\,;\,m_2+m_1,\,2n-1)+(m_1/n-m_1-1)\,F(x\,;\,m_2,n\,;\,m_1-1,n)\end{aligned} \quad .. \quad (2.9)$$

This is obtained by using (2.81) and remembering that here $Q=1$. (2.89) is to be used when $m_2 > m_1$, and (2.9) when $m_2 < m_1$, our object in any case being to get to $F(x;\,m',\,n;\,m',\,n)$ where m' is m_1 or m_2 whichever is less.

As an alternative procedure we might also have had by using (2.83)

$$\begin{aligned}F(x\,;\,m_2,n\,;\,m_1,n)=&\;(1/m_2+1)\,F_0(m_2+1,n-1;\,x)\,F(x;\,m_1,n)-(1/m_2+1)\,F(x;\,m_2+1+m_1,2n-1)\\&+(n-m_2-2/m_2+2)\,F(x\,;\,m_2+1,\,n\,;\,m_1,n)\end{aligned} \quad .. \quad (2.91)$$

or

$$\begin{aligned}F(x\,;\,m_2,n\,;\,m_1,n)=&\;(1/m_1+1)\,F(x;\,m_1+1+m_2,\,2n-1)\\&+(n-m_1-2/m_1+1)\,F(x\,;\,m_2,n\,;\,m_1+1,n)\end{aligned} \quad .. \quad (2.92)$$

We shall use (2.91) or (2.92) according as $m_1 > m_2$ or $m_2 > m_1$ our object there being in any case to get to $F(x\,;\,m'',\,n\,;\,m'',\,n)$ where m'' is m_1 or m_2 whichever is greater.

For $p = 2$, using (2.89) or (2.9), we have (for the distribution of the maximum statistic).

$$F\{x : m+1,\,n\,;\,m,n\} \equiv F\left\{x\,;\begin{pmatrix} m+1 & m\\ m+1 & m\end{pmatrix};n\right\}$$

$$\begin{aligned}&=+F(x\,;\,m+1,n\,;\,m,n)-F(x\,;\,m,n\,;\,m+1,n)\\&=-(1/n-m-2)\,F_0(m+1,n-1\,;\,x)\,F(x\,;\,m,n)+(1/n-m-2)\,F(x\,;\,2m+1,2n-1)\\&\quad+(m+1/n-m-2)\,F(x;\,m,n;\,m,n)+(1/n-m-2)\,F(x\,;\,2m+1,2n-1)\\&\quad-(m+1/n-m-2)\,F(x\,;\,m,n\,;\,m,n)=(2/n-m-2)\,F(x;\,2m+1,2n-1)\\&\quad-(1/n-m-2)\,F_0(m+1,n-1\,;\,x)\,F(x;\,m,n)\end{aligned} \quad .. \quad (2.93)$$

Alternatively by using (2.91) and (2.92) we could have

$$\begin{aligned}&F\left\{x\,;\begin{pmatrix} m+1 & m\\ m+1 & m\end{pmatrix};n\right\}\\&=(2/m+1)F(x;\,2m+1,\,2n-1)-(1/m+1)F_0(m+1,n-1;\,x)\,F(x;\,m+1,n)\end{aligned} \quad .. \quad (2.94)$$

With a general m_2 and m_1 (but $m_2>m_1$) we should have had in place of (2.93) and (2.94)

$$\mathrm{F}\{\mathrm{x}; m_2, n; m_1, n\} \equiv \mathrm{F}\left\{\mathrm{x}; \begin{pmatrix} m_2 & m_1 \\ m_2 & m_1 \end{pmatrix}; n\right\}$$

$$= (2/n-m_2+1)\ \mathrm{F}(\mathrm{x}\ ; m_2+m_1, 2n-1)-(1/n-m_2-1)\ \mathrm{F}_0(m_2, n-1; \mathrm{x})\ \mathrm{F}(\mathrm{x}\ ; m_1, n)$$

$$+(m_2/n-m_2-1)\quad \mathrm{F}\left\{\mathrm{x}; \begin{pmatrix} m_2-1 & m_1 \\ m_2-1 & m_1 \end{pmatrix}; n\right\} \qquad .. \quad (2.95)$$

or alternatively

$$= (2/m_2+1)\ \mathrm{F}(\mathrm{x}\ ; m_1+m_2, 2n-1)-(1/m_1+1)\ \mathrm{F}(m_1+1, n-1\ ; \mathrm{x})\ \mathrm{F}_0(x\ ; m_2, n)$$

$$+(n-m_1-2/m_1+1)\ \mathrm{F}\left\{\mathrm{x}\ ; \begin{pmatrix} m_2 & m_1+1 \\ m_2 & m_1+1 \end{pmatrix}; n\right\} \qquad .. \quad (2.96)$$

(2.95) or (2.96) gives us a recursion formula by which we get after a few steps to

$$\mathrm{F}\left\{\mathrm{x}\ ; \begin{pmatrix} m_1 & m_1 \\ m_1 & m_1 \end{pmatrix}; n\right\} \textit{ or } \mathrm{F}\left\{\mathrm{x}\ ; \begin{pmatrix} m_2 & m_2 \\ m_2 & m_2 \end{pmatrix}; n\right\} \qquad .. \quad (2.961)$$

either of which is zero by definition of the functions F{ }.

This enables us to evaluate $\mathrm{F}\{\mathrm{x}\ ; m_2, n\ ; m_1, n\}$ in terms of functions lik $\mathrm{F}(x\ ; m', n')$ which are readily evaluated from the incomplete B-function tables and functions like $\mathrm{F}_0(\mathrm{x}\ ; m', n')$ which are of course directly calculated.

Similarly we have by using (2.84)

$$\mathrm{F}(m_2, n; m_1, n; \mathrm{x}) \equiv \int_{x_1}^{\infty} \frac{x_2^{m_2}\ dx_2}{(1+x_2)^n} \int_{\mathrm{x}}^{\infty} \frac{x_1^{m_1}\ dx_1}{(1+x_1)^n}$$

$$= (1/n-m_2-1)\ \mathrm{F}(m_2+m_1, 2n-1; \mathrm{x})+(m_2/n-m_2-1)\ \mathrm{F}(m_2-1, n{:}\ m_1\ n{:}\ \mathrm{x}) \qquad .. \quad (2.97)$$

or on the other hand

$$= (1/n-m_1-1)\ \mathrm{F}_0(m_2, n-1\ ; \mathrm{x})\ \mathrm{F}(m_2, n\ ; \mathrm{x})-(1/n-m_1-1)\ \mathrm{F}(m_1+m_2, 2n-1\ ; \mathrm{x})$$

$$+(m_1/n-m_1-1)\ \mathrm{F}(m_2, n\ ; m_1-1, n\ ; \mathrm{x}) \qquad .. \quad (2.98)$$

The form (2.97) (obtained by integration by parts over x_2) or (2.98) (obtained by integration by parts over x_1) is to be used according as $m_2>m_1$ or $m_1>m_2$ our object being in any case to get to $\mathrm{F}(m'\ ; n\ ; m', n\ ; \mathrm{x})$ where m' is the lesser of m_2 and m_1

Alternatively by using (2.85)

$$\mathrm{F}(m_2, n\ ; m_1, n\ ; \mathrm{x}) = -(1/m_2+1)\ \mathrm{F}(m_2+m_1+1, 2n-1\ ; \mathrm{x})$$

$$+(n-m_2-2/m_2+1)\ \mathrm{F}(m_2+1, n\ ; m_1, n\ ; \mathrm{x})$$

$$= -(1/m_1+1)\ \mathrm{F}_0(m_1+1, n-1\ ; \mathrm{x})\mathrm{F}(m_2, n\ ; \mathrm{x})+(1/m_1+1)\ \mathrm{F}(m_1+m_2+1, 2n-1\ ; \mathrm{x})$$

$$+(n-m_2-2/m_1+1)\ \mathrm{F}(m_2, n\ ; m_1+1, n\ ; \mathrm{x}) \qquad .. \quad (2.99)$$

Of the two forms given by (2.99) the first or the second is to be used according as $m_2 > m_1$ or $m_1 > m_2$ our object in any case being to get to $F(m'', n ; m'', n ; x)$ where m'' is the greater of m_2 and m_1. Using (2.97) and (2.98) we easily see (with $m_2 > m_1$) that

$$F\{m_2, n : m_1, n ; x\} \equiv F\left\{\begin{pmatrix} m_2 & m_1 \\ m_2 & m_1 \end{pmatrix} ; n ; x \right\}$$

$$= (2/n-m_2-1)\ F(m_2+m_1, 2n-1 ; x) - (1/n-m_2-1)\ F_0(m_2, n-1 ; x)F(m_1, n ; x)$$

$$+ (m_2/n-m_2-1)\ F\left\{\begin{pmatrix} m_2-1 & m_1 \\ m_2-1 & m_1 \end{pmatrix} ; n ; x\right\} \quad .. \quad (2.991)$$

and using (2.99) we have

$$F\left\{\begin{pmatrix} m_2 & m_1 \\ m_2 & m_1 \end{pmatrix} ; n ; x\right\} = +(2/m_1+1)\ F(m_2+m_1+1, 2n-1 ; x)$$

$$-(1/m_1+1)\ F_0(m_1+1, n-1 ; x)\ F(m_2, n ; x)$$

$$+(n-m_1-2/m_1+1)\ F\left\{\begin{pmatrix} m_2 & m_1+1 \\ m_2 & m_1+1 \end{pmatrix} ; n ; x\right\} \quad .. \quad (2.992)$$

By using (2.991) or (2.992) we can by successive stages get to

$$F\left\{\begin{pmatrix} m_1 & m_1 \\ m_1 & m_1 \end{pmatrix} ; n ; x\right\} \quad \text{or} \quad F\left\{\begin{pmatrix} m_2 & m_2 \\ m_2 & m_2 \end{pmatrix} ; n ; x\right\}$$

either of which is zero by definition. This means that we can express $F\{m_2, n ; m_1, n ; x\}$ in terms of functions of the form $F(m', n' ; x)$ which are calculable in terms of the incomplete B-functions tables. For the distribution of the minimum statistic (for $p = 2$)

$$F\{m+1, n ; m, n ; x\} \equiv F\left\{\begin{pmatrix} m+1 & m \\ m+1 & m \end{pmatrix} ; n ; x\right\}$$

$$\left.\begin{aligned} &= (2/n-m-2)\ F(2m+1, 2n-1 ; x) - (1/n-m-2)\ F_0(m+1, n-1 ; x)F(m, n ; x) \\ &= (2/m+1)\ F(2m+2, 2n-1 ; x) - (1/m+1)\ F_0(m+1, n-1 ; x)\ F(m+1, n ; x) \end{aligned}\right\} \quad (2{\cdot}993)$$

Except for a constant factor (2.93) and (2.94) give us two alternative forms for the distribution of the maximum statistic for $p = 2$, and (2.993) gives us two alternative forms for the distribution of the minimum statistic for the case $p = 2$.

3. THE ACTUAL DERIVATION OF THE DISTRIBUTION OF THE MAXIMUM AND THE MINIMUM STATISTIC FOR THE GENERAL CASE.

Before proceeding with the actual derivation for the general case we shall prove a few more auxiliary theorems or lemmas.

It follows from the notation and definition discussed in §2 that whether in $F\{x\,;(m_p, m_{p-1}, \ldots m_1)\,;\,n\}$ or $F\{(m_p, m_{p-1}, \ldots m_1)\,;\,n\,;\,x\}$ that is, in

$$F\left\{x\,;\begin{pmatrix} m_p & m_{p-1} & .. & m_1 \\ m_p & m_{p-1} & .. & m_1 \\ . & . & .. & . \\ m_p & m_{p-1} & .. & m_1 \end{pmatrix};n\right\} \quad \text{or} \quad F\left\{\begin{pmatrix} m_p & m_{p-1} & .. & m_1 \\ m_p & m_{p-1} & .. & m_1 \\ . & . & .. & . \\ m_p & m_{p-1} & .. & m_1 \end{pmatrix};n\,;\,x\right\}$$

if we put $m_s = m_{s'}$, ($s \neq s'$, and $s, s' = p, p-1 \ldots 1$) then $F\{\quad\}$ becomes zero. Stated otherwise, if in the pseudo-determinantal forms we put any two columns equal then $F\{\quad\}$ becomes zero. This is evident if we look at the expansions (2.76) or (2.77).

Turning now to functions $F(\quad)$ we notice first that

$$\begin{aligned} & F(x\,;m_2,n\,;m_1,n)+F(x\,;m_1,n\,;m_2,n) \\ & = \int_0^x \frac{x_2^{m_2}\,dx_2}{(1+x_2)^n}\int_0^{x_2}\frac{x_1^{m_1}\,dx_1}{(1+x_1)^n}+\int_0^x \frac{x_2^{m_1}\,dx_2}{(1+x_2)^n}\int_0^{x_2}\frac{x_1^{m_2}\,dx_1}{(1+x_1)^n} \\ & = F(x\,;m_2,n)F(x\,;m_1,n) \qquad .. \quad (3.1) \end{aligned}$$

Likewise, if in $F(x\,;m_p,n\,;m_{p-1},n\,;\ldots m_1,n)$ we permute the parameters $m_p, m_{p-1}, \ldots m_1$ over the different places then

$$\Sigma F(x\,;m'_p,n\,;m'_{p-1},n;\,..\,m'_1,n) = \prod_{s=1}^{p} F(x\,;m_s,n) \qquad .. \quad (3.11)$$

where $(m'_p, m'_{p-1}, \ldots, m'_1)$ is any permutation of $(m_p, m_{p-1}, \ldots, m_1)$ and where the summation Σ extends over all such possible permutations.

Similarly we have also

$$\Sigma F(m'_p,n\,;m'_{p-1},n\,;\,..\,m'_1,n\,;x) = \prod_{s=1}^{p} F(m_s,n\,;x) \qquad .. \quad (3.12)$$

Going back to the expansions (2.74) and (2.75) we notice that we could have generalised

$$F\left\{x;\begin{pmatrix} m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ . & . & .. & . \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \end{pmatrix}\right\} \text{and } F\left\{\begin{pmatrix} m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ . & . & .. & . \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \end{pmatrix};x\right\}$$

to

$$F\left\{x:\begin{pmatrix} m_{pp},n_{pp} & m_{p,p-1},n_{p,p-1} & .. & m_{p1},n_{p1} \\ . & . & .. & . \\ m_{1p},n_{1p} & m_{1,p-1},n_{1,p-1} & .. & m_{11},n_{11} \end{pmatrix}\right\} \text{and } F\left\{\begin{pmatrix} m_{pp},n_{pp} & m_{p,p-1},n_{p,p-1} & .. & m_{p1},n_{p1} \\ . & . & .. & . \\ m_{1p},n_{1p} & m_{1,p-1},n_{1,p-1} & .. & m_{11},n_{11} \end{pmatrix}:x\right\}$$

or to $\quad F\{x : (m_{ij}, n_{ij})\}$ and $F\{(m_{ij}, n_{ij})\,;x\}$ .. $(i, j=p, p-1, \,..\, 1)$ .. (3.13)

The special case we have considered earlier is obtained if we put $m_{ij} = m_{oj}$, $n_{ij} = n_{oj}$ $(j=p, \ldots, 1)$ and the more special case directly arising out of the distribution problem to be considered here is obtained if further we put $m_{oj} = m_o + j$, $n_{oj} = n$ $(j = p, p-1, \ldots, 1)$. The law of expansion of (3.13) is, of course, evident.

Of great use for later investigations will be an appropriate notation which covers the case of all other rows of the pseudo-determinants being like $(m_p, n_p\,;\,m_{p-1}, n_{p-1}, \ldots m_1, n_1)$.

and only one row being like $(m'_p, n'_p\,;\, m'_{p-1}, n'_{p-1}; \ldots m'_1, n'_1)$. Where $(m'_p, n'_p\,;\, m_{p-1}', n'_{p-1}\,; \ldots m'_1, n'_1)$ occurs in the s-th row we shall change our symbols a little and write for (3.13)

$$F\{x\,;\,(m^s_{ij}, n^s_{ij})\} \text{ and } F\{(m^s_{ij}, n^s_{ij});\,x\} \qquad .. \quad (3.14)$$

$$\left.\begin{array}{ll} \text{where} & m^s_{ij}=m_{oj},\ n^s_{ij}=n_{oj}\ (i\neq s \text{ and } j=p, p-1, \,..\, 1) \\ \text{and} & m^s_{ij}=m'_j,\ n^s_{ij}=n'_j(i=s, \text{ and } j=p,\ p-1, \,..\, 1) \end{array}\right\} \qquad .. \quad (3.15)$$

For the more special use arising directly out of the distribution problem we have, of course,

$$\left.\begin{array}{l} m^s_{ij}=m_o+j\ ;\ n^s_{ij}=n(i\neq s, \text{ and } j=p-1, \,..\, 0) \\ m^s_{ij}=m'_j\ ;\ n^s_{ij}=n'_j(i=s, \text{ and } j=p-1, \,..\, 0) \end{array}\right\} \qquad .. \quad (3.16)$$

It will be seen that instead of the single suffix notation $(m_p, n_p\,;\, m_{p-1}, n_{p-1}\,; \ldots m_1, n_1)$ we use for convenience the two-way and double suffix notation

$$(m_{op},\ n_{op}\ ;\ m_{o,p-1}\ ;\ n_{o,p-1}, \ldots m_{o1}, n_{o1})$$

Pursuing the line suggested by (3.1), (3.11) and (3.12) we have now

$$F\left\{x\,;\begin{pmatrix} m'_p,n'_p & m'_{p-1},n'_{p-1} & .. & m'_1,n'_1 \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ . & . & .. & . \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \end{pmatrix}\right\} - F\left\{x\,;\begin{pmatrix} m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ m'_p,n'_p & m'_{p-1},n'_{p-1} & .. & m'_1,n'_1 \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ . & . & .. & . \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \end{pmatrix}\right\}$$

$$+F\left\{x\,;\begin{pmatrix} m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ m'_p,n'_p & m'_{p-1},n'_{p-1} & .. & m'_1,n'_1 \\ . & . & .. & . \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \end{pmatrix}\right\} - F\left\{ \qquad \right\} + F\left\{ \qquad \right\} - \ldots$$

$$+(-1)^{p-1}\,F\left\{x\,;\begin{pmatrix} m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ m_p,n_p & m_{p-1},n_{p-1} & .. & m_1,n_1 \\ . & . & .. & . \\ m'_p,n'_p & m'_{p-1},n'_{p-1} & .. & m'_1,n'_1 \end{pmatrix}\right\}$$

$$=F(x\,;m'_p,n'_p)\,F\left\{x\,;\begin{pmatrix} m_{p-1},n_{p-1} & .. & m_1,n_1 \\ m_{p-1},n_{p-1} & .. & m_1,n_1 \\ . & .. & . \\ m_{p-1},n_{p-1} & .. & m_1,n_1 \end{pmatrix}\right\} - F(x;m'_{p-1},n'_{p-1})F\left\{x\,;\begin{pmatrix} m_p,n_p & m_{p-2},n_{p-2} & .. & m_1,n_1 \\ m_p,n_p & m_{p-2},n_{p-2} & .. & m_1,n_1 \\ . & . & .. & . \\ m_p,n_p & m_{p-2},n_{p-2} & .. & m_1,n_1 \end{pmatrix}\right\}$$

$$+\ldots+(-1)^{p-1}F(x\,;\,m'_1,n'_1)\ F\left\{x\,;\begin{pmatrix} m_p,n_p & .. & m_2,n_2 \\ m_p,n_p & .. & m_2,n_2 \\ . & .. & . \\ m_p,n_p & .. & m_2,n_2 \end{pmatrix}\right\} \qquad .. \quad (3\cdot 17)$$

or in the more compact notation of (3.14) and (3.15)

$$\sum_{s=1}^{p-1} (-1)^{p-s}\quad F\{x\,;\,(m^s_{ij},n^s_{ij})\}$$

$$=\sum_{j=p}^{1} (-1)^{p-1}F(x\,;\,m'_j,n'_j)\ F\left\{x\,;\begin{pmatrix} m_{op},n_{op} & .. & m_{o,j+1},\ n_{o,j+1} & m_{o,j-1},n_{o,j-1} & .. & m_{o1},n_{o1} \\ m_{op},n_{op} & .. & m_{o,j+1},\ n_{o,j+1} & m_{o,j-1},n_{o,j-1} & .. & m_{o1},n_{o1} \\ . & .. & . & . & .. & . \\ m_{op},n_{op} & .. & m_{o,j+1},\ n_{o,j+1} & m_{o,j-1},n_{o,j-1} & .. & m_{o1},n_{o1} \end{pmatrix}\right\} \quad (3.18)$$

where m^s_{1j} and n^s_{1j} are given by (3.15) and $m_{1j} = m_{oj}, n_{1j} = n_{oj}$ $(j = p, p-1, \ldots, 1)$; we could if we like drop the 'o' of the double suffix throughout; this will be done in some of the investigations to follow, specially in cases where no confusion is likely to be caused thereby.

This summation $\sum\limits_{j=1}^{p}$ carries obvious limitations at the extremities. We should have likewise

$$\sum_{s=0}^{p-1} (-1)^{p-s} \, F\{(m^s_{ij}, n^s_{ij}) \,;\, x\} = \sum_{j=0}^{p-1} (-1)^{p-s} \, F(m'_j, n'_j ; x) \; F\{Co(m_{pj}, n_{pj}) ;\, x\} \quad .. \quad (3.19)$$

The mechanism by which (3.17), that is, (3.18) or (3.19) is obtained may be illustrated by considering in particular the case of $p = 3$ and working it out in detail. Here we have

$$F\left\{x \,;\begin{pmatrix} m'_3,n'_3 & m'_2,n'_2 & m_1',n'_1 \\ m_3, n_3 & m_2, n_2 & m_1, n_1 \\ m_3, n_3 & m_2, n_2 & m_1, n_1 \end{pmatrix}\right\} - F\left\{x \,;\begin{pmatrix} m_3, n_3 & m_2, n_2 & m_1, n_1 \\ m'_3,n'_3 & m'_2,n'_2 & m'_1,n'_1 \\ m_3, n_3 & m_2, n_2 & m_1, n_1 \end{pmatrix}\right\}$$

$$+F\left\{x \,;\begin{pmatrix} m_3, n_3 & m_2, n_2 & m_1, n_1 \\ m_3, n_3 & m_2, n_2 & m_1, n_1 \\ m'_3,n'_3 & m'_2,n'_2 & m'_1,n'_1 \end{pmatrix}\right\}$$

$$= \int_0^x \frac{x_3^{m'_3} dx_3}{(1+x_3)^{n'_3}} \left(\int_0^{x_3} \frac{x_2^{m_2}\, dx_2}{(1+x_2)^{n_2}} \int_0^{x_2} \frac{x_1^{m_1}\, dx_1}{(1+x_1)^{n_1}} - \int_0^{x_3} \frac{x_2^{m_1}\, dx_2}{(1+x_2)^{n_1}} \int_0^{x_2} \frac{x_1^{m_2}\, dx_1}{(1+x_1)^{n_2}} \right)$$

$$+ \left(\int_0^x \frac{x_3^{m_2} dx_3}{(1+x_3)^{n_2}} \int_0^{x_3} \frac{x_2^{m'_3} dx_2}{(1+x_2)^{n'_3}} \int_0^{x_2} \frac{x_1^{m_1}\, dx_1}{(1+x_1)^{n_1}} - \int_0^x \frac{x_3^{m_1} dx_3}{(1+x_3)^{n_1}} \int_0^{x_3} \frac{x_2^{m'_3} dx_2}{(1+x_2)^{n'_3}} \int_0^{x_2} \frac{x_1^{m_2} dx_1}{(1+x_1)^{n_2}} \right)$$

$$+ \left(\int_0^x \frac{x_3^{m_2} dx_3}{(1+x_3)^{n_2}} \int_0^{x_3} \frac{x_2^{m_1}\, dx_2}{(1+x_2)^{n_1}} \int_0^{x_2} \frac{x_1^{m'_3} dx_1}{(1+x_1)^{n'_3}} - \int_0^x \frac{x_3^{m_1} dx_3}{(1+x_3)^{n_1}} \int_0^{x_3} \frac{x_2^{m_2}\, dx_2}{(1+x_2)^{n_2}} \int_0^{x_2} \frac{x_1^{m'_3} dx_1}{(1+x_1)^{n'_3}} \right)$$

$$+ \;(\quad)\; + \;(\quad)\; + \;(\quad)\; \ldots\ldots\ldots\ldots$$

$$= F(x \,;\, m'_3,n'_3) \; F\left\{x \,;\begin{pmatrix} m_2,n_2 & m_1,n_1 \\ m_2,n_2 & m_1,n_1 \end{pmatrix}\right\} - F(x \,;\, m'_2 n'_2) F\left\{x \,;\begin{pmatrix} m_3,n_3 & m_1,n_1 \\ m_3,n_3 & m_1,n_1 \end{pmatrix}\right\}$$

$$+F(x ; m'_1,n'_1) \; F\left\{x \,;\begin{pmatrix} m_3,n_3 & m_2,n_2 \\ m_3,n_3 & m_2,n_2 \end{pmatrix}\right\} \quad .. \quad (3.2)$$

We can at this stage fruitfully take up the main distribution problem.

$$F\left\{x \,;\begin{pmatrix} m_p,n & m_{p-1},n & .. & m_1,n \\ m_p,n & m_{p-1},n & .. & m_1,n \\ \cdot & \cdot & .. & \cdot \\ m_p,n & m_{p-1},n & .. & m_1,n \end{pmatrix}\right\} = \Sigma \pm F(x \,;\, m'_p,n \,;\, m'_{p-1},n \,:\, ..\, m'_1,n)$$

where $(m'_p, m'_{p-1}, \ldots, m'_1)$ is any permutation of the symbols $(m_p, m_{p-1}, \ldots, m_1)$ and the sign + or — is to be taken according as it is an even permutation or an odd permutation. The number of terms in the summation is, of course, $p\,!$. Suppose that $m_p > m_{p-1} > \ldots > m_1$ and suppose further we try to decrease m_p in successive stages till we get to m_{p-1}. Suppose that in any term $F(x \,;\, m'_p, n \,;\, m'_{p-1}, n ; \ldots m'_1, n)$, m'_s is the largest of $(m'_p, m'_{p-1}, \ldots, m'_1)$, that is, $m'_s = m_p$. Then using the fundamental relations (2.81) and (2.82) it is easily seen that

$$\begin{aligned} F(x; m'_p,n; m'_{p-1},n; \,..\, m'_1,n) &= (1/n-m_p-1)\, F(x; m'_p,n; \,..\, m'_{s+1}+m_p, 2n-1; m'_{s-1},n; ..\, m'_1,n) \\ &+ (1/n-m_p-1)\, F(x \,;\, m'_p,n \,;\, ..\, m'_{s+1},n \,;\, m'_{s-1}+m_p,\; 2n-1 \,;\, ..\, m'_1,n) \\ &+ (m_p/n-m_p-1)\, F(x \,;\, m'_p,n \,;\, m'_{p-1},n \,;\, ..\, m'_{s+1},n;\; m_p-1,n \,;\, ..\, m'_1,n) \quad .. \quad (3.21) \end{aligned}$$

he above relations will hold for $s = p-1, p-2, \ldots, 2$

If $s = p$ we shall have instead

$$\begin{aligned} F(x;\, m_p,n;\, m'_{p-1},n;\, \ldots\, m'_1,n) = &-(1/n-m_p-1)\ F_0(x;\, m_p,n-1)\ F(x;\, m'_{p-1},n;\, m'_{p-2},n;\, \ldots\, m'_1,n) \\ &+(1/n-m_p-1)\ F(x;\, m_p+m_{p-1},\, 2n-1\, ;\, m'_{p-2},n;\, \ldots m'_1,n) \\ &+(m_p/n-m_p-1)\ F(x;\, m_p-1,\, n;\, m'_{p-1},n;\, \ldots\, m'_1,n) \qquad \ldots \quad (3.22) \end{aligned}$$

and finally if $s = 1$ we shall have

$$\begin{aligned} F(x;\, m'_p,n;\, m'_{p-1},n;\, \ldots\, m'_2,n;\, m_1,n) = &-(n/n-m_p-1)\ F(x\, ;\, m'_p,n;\, m'_{p-1},n;\, \ldots m'_2+m_p,\, 2n-1) \\ &+(m_p/n-m_p-1)\ F(x\, ;\, m'_p,n\, ;\, m'_{p-1},n\, ;\, \ldots\, m'_2,n\, ;\, m_p-1,\, n) \qquad \ldots \quad (3.23) \end{aligned}$$

The left hand sides of (3.21), (3.22) and (3.23) are all p-fold incomplete integrals, but the first two terms on the right hand side of (3.21), the second term and the second factor of the first term on the right hand side of (3.22) and the first term on the right hand side of (3.23) are all $(p-1)$-fold incomplete integrals. The last terms on the right hand side of (3.21) and (3.23) are all p-fold incomplete integrals again but with one parameter m_p decreased (by one) to m_p-1. As a more convenient notation we can replace

(*i*) $F(x\, ;\, m'_p,n\, ;\, \ldots\, m'_{s+1}+m'_p,\, 2n-1\, ;\, m'_{s-1},n\, ;\, \ldots\, m'_1,n)$ by

$F(x\, ;\, m'_p,n\, ;\, \ldots\, m'_{s+1},n\, ;\, \overleftarrow{m_p},\overleftarrow{n-1}\, ;\, m'_{s-1},n\, ;\, \ldots\, m'_1,n)$

where $\overleftarrow{m_p}$ is always added to the 'm' on the left and $\overleftarrow{n-1}$ to 'n' on the left. $\qquad \ldots \quad (3.24)$

(*ii*) $F_0(x\, ;\, m_p,n-1)\ F(x\, ;\, m'_{p-1},n\, ;\, m'_{p-2},n;\, \ldots\, m'_1,n)$ by

$F(x\, ;\, \overleftarrow{m_p},\overleftarrow{n-1};\, m'_{p-1},n;\, \ldots m'_1,n)$

(*iii*) $F(x\, ;\, m'_p,n\, ;\, \ldots m'_{s+1},n\, ;\, m_p+m_{s-1},\, 2n-1\, ;\, m_{s-2},n\, ;\, \ldots\, m'_1,n)$ by

$F(x\, ;\, m'_p,n\, ;\, \ldots\, m'_{s+1},n\, ;\, \overrightarrow{m_p},\overrightarrow{n-1}\, ;\, m'_{s-1},n\, ;\, \ldots\, m'_1,n)$

where $\overrightarrow{m_p}$ is always added to the 'm' on the right and $\overrightarrow{n-1}$ to the 'n' on the right

Using the results (2.81)—(3.23) and the notation (3.24) we find that

$$\begin{aligned} F\left\{ x\, ;\, \begin{pmatrix} m_p,n & m_{p-1},n & \ldots & m_1,n \\ m_p,n & m_{p-1},n & \ldots & m_1,n \\ \cdot & \cdot & \ldots & \cdot \\ m_p,n & m_{p-1},n & \ldots & m_1,n \end{pmatrix} \right\} \\ = -(1/n-m_p-1)\ F\left\{ x;\, \begin{pmatrix} \overleftarrow{m_p},\overleftarrow{n-1} & m_{p-1},n & \ldots & m_1,n \\ \overleftarrow{m_p},\overleftarrow{n-1} & m_{p-1},n & \ldots & m_1,n \\ \cdot & \cdot & \ldots & \cdot \\ \overleftarrow{m_p},\overleftarrow{n-1} & m_{p-1},n & \ldots & m_1,n \end{pmatrix} \right\} \\ +(1/n-m_p-1)\ F\left\{ x\, ;\, \begin{pmatrix} \overrightarrow{m_p},\, \overrightarrow{n-1} & m_{p-1},n & \ldots & m_1,n \\ \overrightarrow{m_p},\, \overrightarrow{n-1} & m_{p-1},n & \ldots & m_1,n \\ \cdot & \cdot & \ldots & \cdot \\ \square & m_{p-1},n & \ldots & m_1,n \end{pmatrix} \right\} \\ +(m_p/n-m_p-1)F\left\{ x\, ;\, \begin{pmatrix} m_p-1,n & m_{p-1},n & \ldots & m_1,n \\ m_p-1,n & m_{p-1},n & \ldots & m_1,n \\ \cdot & \cdot & \ldots & \cdot \\ m_p-1,n & m_{p-1},n & \ldots & m_1,n \end{pmatrix} \right\} \qquad \ldots \quad (3.25) \end{aligned}$$

where in the second term on the right hand sides of (3.25), $\square$ in the pseudo-determinant merely means that the corresponding terms in the formal expansion are not to be considered at all ; $\square$ is introduced merely to write the pseudo-determinant in a complete form.

If we use now the relations (3.17), (3.18), (3.21)—(3.23) and the notation (3.24), the right hand side of (3.25) would simplify thus :

$$F\left\{x;\begin{pmatrix}\overleftarrow{m_p},\overleftarrow{n-1} & m_{p-1},n & .. & m_1,n\\ \overleftarrow{m_p},\overleftarrow{n-1} & m_{p-1},n & .. & m_1,n\\ . & . & .. & .\\ \overleftarrow{m_p},\overleftarrow{n-1} & m_{p-1},n & .. & m_1,n\end{pmatrix}\right\}$$

$$=F_0(x\,;m,n)\quad F\left\{x\,;\begin{pmatrix}m_{p-1},n & m_{p-2},n & .. & m_1,n\\ m_{p-1},n & m_{p-2},n & .. & m_1,n\\ . & & .. & .\\ m_{p-1},\text{[illegible]} & m_{p-2},n & .. & m_1,n\end{pmatrix}\right\}$$

$$-F\left\{x;\begin{pmatrix}m_{p-1}+m_p,2n-1 & m_{p-2}+m_p,2n-1 & .. & m_1+m_p,2n-1\\ m_{p-1},n & m_{p-2},n & .. & m_1,n\\ . & . & .. & .\\ m_{p-1},n & m_{p-2},n & & m_1,n\end{pmatrix}\right\}$$

$$+\;(\;)\;-\;(\;)\;..$$

$$+F\left\{x\,;\begin{pmatrix}m_{p-1},n & m_{p-2},n & .. & m_1,n\\ m_{p-1},n & m_{p-2},n & .. & m_1,n\\ . & . & .. & .\\ m_{p-1}+m_p,2n-1 & m_{p-2}+m_p,2n-1 & .. & m_1+m_p,2n-1\end{pmatrix}\right\}$$

$$=\qquad F_0(x;\,m_p,n)\quad F\left\{x\,;\begin{pmatrix}m_{p-1},n & m_{p-2},n & .. & m_1,n\\ m_{p-1},n & m_{p-2},n & .. & m_1,n\\ . & . & .. & .\\ m_{p-1},n & m_{p-2},n & .. & m_1,n\end{pmatrix}\right\}$$

$$-F(x\,;m_{p-1}+m_p,2n-1)\;F\left\{x\,;\begin{pmatrix}m_{p-2},n & m_{p-3},n & .. & m_1,n\\ m_{p-2},n & m_{p-3},n & .. & m_1,n\\ . & . & .. & .\end{pmatrix}\right\}$$

$$+F(x\,;m_{p-2}+m_p,\,2n-1)\;F\left\{x\,;\begin{pmatrix}m_{p-1},n & m_{p-3},n & .. & m_1,n\\ m_{p-1},n & m_{p-3},n & .. & m_1,n\\ . & . & .. & .\end{pmatrix}\right\}$$

$$-\;(\;)\;+\;(\;)\;..$$

$$\pm F_0(x\,;m_1+m_p,\,2n-1)F\left\{x\,;\begin{pmatrix}m_{p-1},n & m_{p-2},n & .. & m_2,n\\ m_{p-1},n & m_{p-2},n & .. & m_2,n\\ . & . & .. & .\end{pmatrix}\right\}$$

$$=F_0(x\,;m_p,n)\,F\left\{x;\begin{pmatrix}m_{p-1},n & m_{p-2},n & .. & m_1,n\\ m_{p-1},n & m_{p-2},n & .. & m_1,n\\ . & . & .. & .\end{pmatrix}\right\}$$

$$-\sum_{s=p-1}^{1}(-1)^{p-s-1}\,F(x;\,m_s+m_p,\,2n-1)\;F\left\{x;\begin{pmatrix}m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n\\ m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n\\ . & .. & . & . & .. & .\end{pmatrix}\right\}\;..\quad[3.26]$$

Likewise with obvious limitations at the extremities, which the $\sum\limits_{s=p-1}^{1}$ summations will always be supposed to carry

$$\mathrm{F}\left\{\mathrm{x}\,;\left(\begin{matrix} \overrightarrow{m_p},\overrightarrow{n}-1 & m_{p-1},n & .. & m_1,n \\ \overrightarrow{m_p},\overrightarrow{n}-1 & m_{p-1},n & .. & m_1,n \\ . & . & .. & . \\ \square & m_{p-1},n & .. & m_1,n \end{matrix}\right)\right\}$$

$$= \mathrm{F}\left\{\mathrm{x}\,;\left(\begin{matrix} m_{p-1}+m_p,2n-1 & m_{p-2}+m_p,2n-1 & .. & m_1+m_p,2n-1 \\ m_{p-1},n & m_{p-2},n & .. & m_1,n \\ . & . & .. & . \\ m_{p-1},n & m_{p-1},n & .. & m_1,n \end{matrix}\right)\right\}$$

$$-\mathrm{F}\left\{\mathrm{x};\left(\begin{matrix} m_{p-1},n & m_{p-2},n & .. & m_1,n \\ m_{p-1}+m_p,2n-1 & m_{p-2}+m_p,2n-1 & .. & m_1+m_p,2n-1 \\ m_{p-1},n & m_{p-2},n & .. & m_1,n \\ . & . & .. & \end{matrix}\right)\right\}$$

$$+\ (\ \ )\ -\ (\ \ )\ \ldots\ldots$$

$$\pm\mathrm{F}\left\{\mathrm{x};\left(\begin{matrix} m_{p-1},n & m_{p-2},n & .. & m_1,n \\ m_{p-1},n & m_{p-2},n & .. & m_1,n \\ . & . & .. & . \\ m_{p-1}+m_p,2n-1 & m_{p-2}+m_p,2n-1 & .. & m_1+m_p,2n-1 \end{matrix}\right)\right\}$$

(the last term being + or − according as p is odd or even)

$$= \sum_{s=p-1}^{1}(-1)^{p-s-1}\,\mathrm{F}(\mathrm{x};\ m_s+m_p,2n-1)\mathrm{F}\left\{\mathrm{x}\,;\left(\begin{matrix} m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n \\ m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n \\ . & .. & . & . & .. & . \\ m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n \end{matrix}\right)\right\} \quad (3{\cdot}27)$$

Combining (3.26) and (3.27) we have now (3.26) reducing to

$$\mathrm{F}\left\{\mathrm{x};\left(\begin{matrix} m_p,n & m_{p-1},n & .. & m_1,n \\ m_p,n & m_{p-1},n & .. & m_1,n \\ . & . & .. & . \\ m_p,n & m_{p-1},n & .. & m_1,n \end{matrix}\right)\right\}$$

$$= -(1/n-m_p-1)\ \ \mathrm{F}_0(\mathrm{x}\ ;\ m_p,n-1)\mathrm{F}\ \left\{\mathrm{x};\left(\begin{matrix} m_{p-1},n & m_{p-2},n & .. & m_1,n \\ m_{p-1},n & m_{p-2},n & .. & m_1,n \\ . & . & .. & . \\ m_{p-1},n & m_{p-2},n & .. & m_1\ n \end{matrix}\right)\right\}$$

$$+(2/n-m_p-1)\sum_{s=p-1}^{1}(-1)^{p-s-1}\mathrm{F}(\mathrm{x};\ m_s+m_p,2n-1)\mathrm{F}\left\{\mathrm{x};\left(\begin{matrix} m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n \\ m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n \\ . & .. & . & . & .. & . \\ m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n \end{matrix}\right)\right\}$$

$$+(m_p/n-m_p-1)\ \mathrm{F}\left\{\mathrm{x};\left(\begin{matrix} m_p-1,n & m_{p-1},n & .. & m_1,n \\ . & . & .. & . \\ m_p-1,n & m_{p-1},n & .. & m_1,n \end{matrix}\right)\right\} \quad ..\quad (3{\cdot}28)$$

The left hand side is a p-th order pseudo-determinant. On the right hand side the first term involves a $(p-1)$th order pseudo-determinant, the second group of terms each involves a $(p-2)$-th order pseudo-determinant, the third term is a pseudo-determinant of the p-th order no doubt, but the parameter m_p is replaced by m_p-1. (3.28) provides us with a recursion chain by which m_p is reduced in successive stages to m_{p-1} and as soon as this happens the whole pseudo-determinant becomes zero as we have already observed.

Proceeding along the recursion chain suggested by (3.28) we have

$$F\left\{x\,;\begin{pmatrix} m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_1,n \\ m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_1,n \\ . & . & .. & . \end{pmatrix}\right\}$$

$$=-F\left\{x\,;\begin{pmatrix} m_{\mathrm{p}-1},n & m_{\mathrm{p}-2},n & .. & m_1,n \\ . & . & .. & . \\ m_{\mathrm{p}-1},n & m_{\mathrm{p}-2},n & .. & m_1,n \end{pmatrix}\right\}\sum_{\mathrm{r}=1}^{\mathrm{m_p-m_{p-1}}}\left({}_{\mathrm{m}}P_{\mathrm{r}-1}/{}^{\mathrm{n-m_p-2+r}}P_{\mathrm{r}}\right)\ F_0(x;\,m_{\mathrm{p}}-r+1,n-1)$$

$$+\sum_{\mathrm{s=p-1}}^{1}(-1)^{\mathrm{p-s-1}}\,F\left\{x\,;\begin{pmatrix} m_{\mathrm{p}-1},n & .. & m_{\mathrm{s}+1},n & m_{\mathrm{s}-1},n & .. & m_1,n \\ . & .. & . & . & .. & . \\ m_{\mathrm{p}-1},n & .. & m_{\mathrm{s}+1},n & m_{\mathrm{s}-1},n & .. & m_1,n \end{pmatrix}\right\}\sum_{\mathrm{r}=1}^{\mathrm{m_p-m_{p-1}}}2Q_1$$

where $$Q_1=\left({}^{\mathrm{m_p}}P_{\mathrm{r}-1}/{}^{\mathrm{n-m_p-2+r}}P_{\mathrm{r}}\right)\ F(x\,;\,m_{\mathrm{s}}+m_{\mathrm{p}}-r+1,2n-1) \qquad ..\quad (3{\cdot}29)$$

or in a different notation already considered earlier

$$F\{x\,;\,(m_{\mathrm{p}},n\ ;\ m_{\mathrm{p}-1},n\ ;\ ..\ m_1,n)\}$$

$$=-F\{x\,;(m_{\mathrm{p}-1},n\,;\ m_{\mathrm{p}-2},n;\ ..\ m_1,n)\}\sum_{\mathrm{r}=1}^{\mathrm{m_p-m_{p-1}}}{}^{\mathrm{m_p}}P_{\mathrm{r}-1}/{}^{\mathrm{n-m_p-2+r}}P_{\mathrm{r}}\ F(x\,;\,m_{\mathrm{s}}+m_{\mathrm{p}}-r+1,2n-1)$$

$$+\sum_{\mathrm{s=p-1}}^{1}(-1)^{\mathrm{p-s-1}}\ F\{x;\ (m_{\mathrm{p}-1},n;\ ..\ m_{\mathrm{s}+1},n;\ m_{\mathrm{s}-1},n;\ ..\ m_1,n)\}\sum_{\mathrm{s}=1}^{\mathrm{m_p-m_{p-1}}}2Q_1 \qquad ..\quad (3{\cdot}3)$$

The p-fold pseudo-determinant is thus thrown back on $(p-1)$ and $(p-2)$-fold ones, and these again on $(p-2)$ and $(p-3)$ and so on till we get to 1-fold cases, that is, functions of the type $F(x\,;\,m,n)$ which are easily calculated from the incomplete B—function tables. Now (3.3) is the reduction formula if one attempts to reduce m_{p} to $m_{\mathrm{p}-1}$; on the other hand if one tried to raise m_1 to m_2 then using again the same body of theorems one would have had

$$F\left\{x;\begin{pmatrix} m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_1,n \\ . & . & .. & . \\ m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_1,n \end{pmatrix}\right\}$$

$$=\ (1/m_1+1)\ F\left\{x;\begin{pmatrix} m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n & m_1+\overset{\leftarrow}{1},n-\overset{\leftarrow}{1} \\ m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n & m_1+\overset{\leftarrow}{1},n-\overset{\leftarrow}{1} \\ . & . & .. & . & . \end{pmatrix}\right\}$$

$$+(1/m_1+1)\ F\left\{x\,;\begin{pmatrix} m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n & m_1+\overset{\rightarrow}{1},n-\overset{\rightarrow}{1} \\ . & . & .. & . & . \\ m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n & m_1+\overset{\rightarrow}{1},n-\overset{\rightarrow}{1} \\ m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n & \square \end{pmatrix}\right\}$$

$$+(n-m_1-2/m_1+1)F\left\{x\,;\begin{pmatrix} m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n & \mathrm{m}_1+1,n \\ m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n & m_1+1,n \\ . & . & .. & . & . \end{pmatrix}\right\}$$

$$=\ (-1)^{\mathrm{p}-1}\ (1/m_1+1)\ F_0(x;\,m_1+1,n-1)\ F\left\{x;\begin{pmatrix} m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n \\ . & . & .. & . \\ m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n \end{pmatrix}\right\}$$

$$+(-1)^{\mathrm{p}}\,(2/m_1+1)\sum_{\mathrm{s=p}}^{2}(-1)^{\mathrm{p-s}}\ F(x;\,m_{\mathrm{s}}+m_1+1,\ 2n-1)\ F\left\{x;\begin{pmatrix} m_{\mathrm{p}},n & .. & m_{\mathrm{s}+1},n & m_{\mathrm{s}-1},n & .. & m_2,n \\ . & .. & . & . & .. & . \\ m_{\mathrm{p}},n & .. & m_{\mathrm{s}+1},n & m_{\mathrm{s}-1},n & .. & m_2,n \end{pmatrix}\right\}$$

$$+(n-m_2-2/m_1+1)\ F\left\{x;\begin{pmatrix} m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n & m_1+1,n \\ . & . & .. & . & . \\ m_{\mathrm{p}},n & m_{\mathrm{p}-1},n & .. & m_2,n & m_1+1,n \end{pmatrix}\right\} \qquad ..\quad (3{\cdot}31)$$

$$-(-1)^{p-1}\ F\left\{x;\begin{pmatrix} m_p,n & m_{p-1},n & .. & m_2,n \\ . & . & .. & . \\ m_p,n & m_{p-1},n & .. & m_2,n \end{pmatrix}\right\}\sum_{r=1}^{m_2-m_1}({}_{n-m_1-r}P_{r-1}/{}_{m_1+r}P_r)\ F_0(x;\ m_1+r,n-1)$$

$$+(-1)^p\sum_{s=p}^{2}(1)^{p-s}\ F\left\{x;\begin{pmatrix} m_p,n & m_{s+1},n & m_{s-1},n & .. & m_2,n \\ . & . & . & .. & . \\ m_p,n & m_{s+1},n & m_{s-1},n & .. & m_2,n \end{pmatrix}\right\}\sum_{r=1}^{m_2-m_1}2Q_2 \quad .. \quad (3.32)$$

where $\qquad Q_2=({}_{n-m_1-r}P_{r-1}/{}_{m_1+r}P_1)\ F(x\ ;\ m_1+r+m_s,\ 2n-1)$

and where, as in (3.29), the left hand side is a p-th order pseudo-determinant while on the right hand side the first term involves a $(p-1)$-th and the second involves up to $(p-2)$-th order pseudo-determinants. In a different notation already used (3.32) can be replaced by

$$F\{x\ :\ (m_p,n\ ;\ m_{p-1},n\ ;\ ..\ m_1,n)\}$$

$$=(-1)^{p-1}\ F\{x\ ;\ (m_p,n\ ;..m_2,u)\}\ \times\sum_{r=1}^{m_1-m_2}({}_{n-m_1-r}P_r/{}_{m_1+r}P_1)\ F_0(x\ ;\ (m_1+r,\ n-1)$$

$$+(-1)^p\sum_{s=p}^{2}(-1)^{p-s}\ F\{x\ ;\ (m_p,n\ ;\ ..\ m_{s+1},n\ ;\ m_{s-1},n\ ;\ ..\ m_1,n)\}\quad\sum_{r=1}^{m_2-m_1}2Q_2$$

where $\qquad Q_2=({}_{n-m_1-r}P_{r-1}/{}_{m_1-r}P_{r-1})\ F\{x\ ;\ (m_s+r+m_1,\ 2n-1)\} \qquad .. \quad (3{\cdot}33)$

In the actual distribution problem of maximum statistic for p variates we have

$$m_j=m+j+1\ \ (j=p,\ p-1,\ \ldots,\ 1)$$

Hence (3.3) is to be replaced by

$$F\{x;\ (m+p-1,\ n\ ;\ m+p-2,\ n\ ;\ ..\ m,n)\}$$

$$=-F\{x\ ;\ (m+p-2,\ n;\ m+p-3,\ ;\ ..\ m,n)\}\quad F_0(x\ ;\ m+p-1,\ n-1)/n-m-p$$

$$+(2/n-m-1)\sum_{s=p-1}^{1}(-1)\ F\{x\ ;\ (m+p-2,n\ ..\ m+s,n\ ;\ m+s-2,\ n;\ ..\ m,n)\}\ Q_3$$

where $\qquad Q_3=F(x\ ;\ 2m+p+s-2,\ 2n-1) \qquad .. \quad 3.34)$

and (3·33) is replaced by

$$F\{x\ ;\ (m+p-1,\ n\ ;\ m+p-2,\ n\ ;\ ..\ m,n)\}$$

$$=\quad(-1)^{p-1}\ F\{x\ ;\ (m+p-1,\ n\ :\ m+p-2,\ n\ ;\ ..\ m+1,n)\ (1/m+1)\ F_0(x\ ;\ m+1,\ n-1)$$

$$+(-1)^p\sum_{s=p}^{2}(-1)^{p-s}\ F\{x\ ;\ (m+p-1,\ n\ ;\ ..\ m+s,\ n\ ;\ m+s-2,\ n\ ;\ ..\ m+1,\ n)\}\ Q_4$$

where $\qquad Q_4=(2/m+1)\ F(x\ ;\ 2m+s,\ 2n-1) \qquad .. \quad (3.35)$

This is so far as the distribution of the maximum of p-statistics is concerned. We shall now turn to the distribution of the minimum statistic.

Remembering (2.84) and (2.85) suppose we try to expand

$$F\{(m_p,n\ ;\ m_{p-1},n\ ;\ ..\ m_1,n)\ ;\ x\}=F\left\{\begin{pmatrix} m_p,n & m_{p-1},n & .. & m_1,n \\ . & . & .. & . \\ m_p,n & m_{p-1},n & .. & m_1,n \end{pmatrix};x\right\}$$

with the proviso that $m_p>m_{p-1}>\ \ldots\ >m_1$

As indicated earlier we can write this as $\Sigma\ F(m'_p,\ n\ ;\ m'_{p-1},\ n\ ;\ \ldots.\ m'_1,\ n\ ;\ x)$ where $(m'_p,\ m'_{p-1},\ \ldots,\ m'_1)$ is any permutation of $(m_p,\ m_{p-1},\ \ldots,\ m_1)$ and the summation is over all such permutations.

Consider the typical term $F(m'_{p}, n\,;\, m'_{p-1}, n\,;\, \ldots m'_{s}, n\,;\, m'_{s-1}, n\,;\, \ldots;\, m'_{1}, n\,;\, x)$ and suppose that $m'_{s} = m_{p}$. Suppose further that we try to reduce m_{p} in successive stages to m_{p-1}. Then using (2.84) we have

$$
\begin{aligned}
& F(m'_{p},n\,;\, ..\, m'_{s+1},n\,;\, m_{p},n\,;\, m'_{s-1},n\,;\, ..\, m'_{1},n\,;\, x) \\
&= (1/n-m_{p}-1)\, F(m'_{p},n\,;\, ..\, m'_{s+1},n\,;\, m_{p}+m'_{s-1},\, 2n-1\,;\, m'_{s-2}, n\,;\, ..\, m'_{1},n\,;\, x) \\
&\quad -(1/n-m_{p}-1)\, F(m'_{p},n\; ..\, m'_{s+1}+m_{p},\, 2n-1\,;\, m'_{s-1},n\,;\, ..\, m'_{1},n\,;\, x) \\
&\quad +(m_{p}/n-m_{p}-1)\, F(m'_{p},n\,;\, ..\, m'_{s+1},n\,;\, m_{p}-1, n\,;\, m'_{s-1}, n\,;\, ..\, m'_{1}, n;\, x) \quad .. \quad (3.36)
\end{aligned}
$$

When $s = p$, the middle term would not be there, and when $s=1$, the first term is to be replaced by $F_0(m_{p}, n-1\,;\, x)\; F(m'_{p}, n\,;\, m'_{p-1}, n\,;\, \ldots m'_{2}, n\,;\, x)/(n-m_{p}-1)$. As earlier introduce now the notation

$$F(m'_{p},n\,;\; ..\, m'_{s+1},n\,;\, \overset{\rightarrow}{m_{p}},\vec{n}-1\,;\, m'_{s-1},n\,;\, ..\, m'_{1},n\,;\, x)$$

and $F(m'_{p},n\,;\, ..\, m'_{s+1},n\,;\, \overset{\leftarrow}{m_{p}},\overset{\leftarrow}{n}-1\,;\, m'_{s-1},n\,;\, ..\, m'_{1},n\,;\, x)$ for

$F(m'_{p},n\,;\, ..\, m'_{s+1},n\,;\, m'_{s-1}+m_{p},\, 2n-1\,;\, m'_{s-2},n\,;\, ..\, m'_{1},n\,;\, x)$ and

$F(m'_{p},n;\; ..\, m'_{s+1}+m_{p},\, 2n-1;\, m'_{s-1},n;\, ..\, m'_{1},n;\, x)$ respectively, with the proviso that $F(\overset{\leftarrow}{m_{p}},\, \overset{\leftarrow}{n}-1;\, m'_{p-1},n;\, ..\, m'_{1},n';\, x)$ would denote $F_0(m_{p},n-1;\, x)\; F(m'_{p-1},n;\, ..\, m'_{1},n;\, x)$.
Then we can replace (3,36) by

$$
\begin{aligned}
& F(m'_{p},n\,;\, ..\, m'_{s+1},n\,;\, m_{p},n\,;\, m'_{s-1}, n;\, ..\, m'_{1},n\,;\, x) \\
&\quad = (1/n-m_{p}-1)\, F(m'_{p},n\,;\, ..\, m'_{s+1},n\,;\, \overset{\rightarrow}{m_{p}},\vec{n}-1\,;\, m'_{s-1},n\,;\, ..\, m'_{1},n\,;\, x) \\
&\quad -(1/n-m_{p}-1)\, F(m'_{p},n\,;\, ..\, m'_{s+1},n\,;\, \overset{\leftarrow}{m_{p}},\overset{\leftarrow}{n}-1\,;\, m'_{s-1},n\,;\, ..\, m'_{1},n\,;\, x) \\
&\quad +(m_{p}/n-m_{p}-1)\, F(m'_{p},n\,;\, ..\, m'_{s+1},n\,;\, m_{p}-1,n\,;\, m'_{s-1},n\;\; ..\, m'_{1},n\,;\, x)
\end{aligned}
$$

Hence we have

$$
\begin{aligned}
& F\left\{\begin{pmatrix} m_{p},n & m_{p-1},n & .. & m_{1},n \\ \cdot & \cdot & .. & \cdot \\ m_{p},n & m_{p-1},n & .. & m_{1},n \end{pmatrix};\, x\right\} \\
= \;& (1/n-m_{p}-1)\; F\left\{\begin{pmatrix} \square & m_{p-1},n & .. & m_{1},n \\ \overset{\rightarrow}{m_{p}},\vec{n}-1 & m_{p-1},n & .. & m_{1},n \\ \cdot & \cdot & .. & \cdot \\ \overset{\rightarrow}{m_{p}},\vec{n}-1 & m_{p-1},n & .. & m_{1},n \end{pmatrix}:\, x\right\} \\
& -(1/n-m_{p}-1)\; F\left\{\begin{pmatrix} \overset{\leftarrow}{m_{p}},\overset{\leftarrow}{n}-1 & m_{p-1},n & .. & m_{1},n \\ \overset{\leftarrow}{m_{p}},\overset{\leftarrow}{n}-1 & m_{p-1},n & .. & m_{1},n \\ \cdot & \cdot & .. & \cdot \\ \overset{\leftarrow}{m_{p}},\overset{\leftarrow}{n}-1 & m_{p-1},n & .. & m_{1},n \end{pmatrix};\, x\right\} \\
& +(m_{p}/n-m_{p}-1)\; F\left\{\begin{pmatrix} m_{p}-1,n & m_{p-1},n & .. & m_{1},n \\ m_{p}-1,n & m_{p-1},n & .. & m_{1},n \\ \cdot & \cdot & .. & \cdot \\ m_{p}-1,n & m_{p-1},n & .. & m_{1},n \end{pmatrix};\, x\right\} \qquad .. \quad (3.37)
\end{aligned}
$$

Expanding and proceeding exactly as for the case of the maximum statistic we have

$$
\begin{aligned}
&F\{(m_p,n\,;\,m_{p-1},n\,;\,..\,m_1,n)\,;\,x\}\\
&=-(1/n-m_p-1)F_0(m_p,\,n-1\,;\,x)\times F\{(m_{p-1},n\,;\,m_{p-2},\,n\,;\,..\,m_1,n)\,;\,x\}\\
&+(2/n-m_p-1)\sum_{s=p-1}^{1}F(m_s+m_p,\,2n-1\,;\,x)\,Q_5\\
&+(m_p/n-m_p-1)\,F\left\{\begin{pmatrix} m_p-1,n & m_{p-1},n & .. & m_1,n\\ m_p-1,n & m_{p-1},n & .. & m_1,n\\ . & . & .. & .\\ m_p-1,n & m_{p-1},n & .. & m_1,n\end{pmatrix};x\right\} \qquad ..\quad (3.38)
\end{aligned}
$$

where $$Q_5=F\left\{\begin{pmatrix} m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n\\ . & & . & . & & .\\ m_{p-1},n & .. & m_{s+1},n & m_{s-1},n & .. & m_1,n\end{pmatrix};x\right\}$$

We have thus by successive stages as before

$$
\begin{aligned}
&F\{(m_p,n\,;\,..\,m_1,n)\,;\,x\}\\
&=-F\{(m_{p-1},n\,;\,m_{p-2},n\,;\,..\,m_1,n)\,;\,x\}\times\sum_{r=1}^{m_p-m_{p-1}}({}_{m_p}P_{r-1}/{}_{n-m_p-1}P_r)\,F_0(m_p-r+1,\,n-1\,;\,x)\\
&+2\sum_{s=p-1}^{1}(-1)^{p-s-1}F\{(m_{p-1},n\,;\,..\,m_{s+1},n\,;\,m_{s-1},n\,;\,..\,m_1,n)\,;\,x\}\sum_{r=1}^{m-_pm_{p-1}}Q_6
\end{aligned}
$$

where $$Q_6=({}_{m_p}P_{r-1}/{}_{n-m_p-r}P_r)\,F(m_s+m_p-r+1,\,2n-1\,;\,x) \qquad ..\quad (3.39)$$

This is if we reduce m_p by successive stages to m_{p-1} ; on the other hand if we try to increase m_1 by successive stages to m_2 then proceeding exactly as for the case of the maximum statistic we have, by using (2.85),

$$
\begin{aligned}
F\{(m_p,n\,;\,..\,m_1,n)\,;\,x\}=&-(1/m_1+1)\,F\left\{\begin{pmatrix} m_p,n & m_{p-1},n & .. & m_2,n & \square\\ m_p,n & m_{p-1},n & .. & m_2,n & \overrightarrow{m_1+1},\overrightarrow{n-1}\\ . & . & .. & . & .\\ m_p,n & m_{p-1},n & .. & m_2,n & \overrightarrow{m_1+1},\overrightarrow{n-1}\end{pmatrix}\right\}\\
&+(1/m_1+1)\,F\left\{\begin{pmatrix} m_p,n & m_{p-1},n & .. & m_2,n & \overleftarrow{m_1+1},\ \overleftarrow{n-1}\\ m_p,n & m_{p-1},n & .. & m_2,n & \overleftarrow{m_1+1},\ \overleftarrow{n-1}\\ . & . & .. & . & .\\ m_p,n & m_{p-1},n & .. & m_2,n & \overleftarrow{m_1+1},\ \overleftarrow{n-1}\end{pmatrix};x\right\}\\
&+(n-m_1-2/m_1+1)\,F\left\{\begin{pmatrix} m_p,n & .. & m_2,n & m_1+1,n\\ . & .. & . & .\\ m_p,n & .. & m_2,n & m_1+1,n\end{pmatrix};x\right\} \quad ..\ (3.4)
\end{aligned}
$$

whence we have

$$
\begin{aligned}
&F\{(m_p,n\;..\;m_1,n);\,x\}\\
&=\ (-1)^{p-1}\,F\{m_p,n;\,m_{p-1},n;\,..\,m_2,n);\,x\}\times\sum_{r=1}^{m_2-m_1}({}_{n-m_1-r}P_r/{}_{m_1-r}P_r)\,F_0(m_p+r,n-1)\\
&\quad-\sum_{s=p}^{2}(-1)^{p-s}\,F\{m_p,n;\,..\,m_{s+1},n;\,m_{s-1},n;\,..\,m_2,n);\,x\}\sum_{r=1}^{m_2-m_1}2Q_7
\end{aligned}
$$

where $$Q_7=({}_{n-m_1-r}P_{r-1}/{}_{m_1+r}P_r)\,F(m_1+r+m_s,\,2n-1;\,x) \qquad ..\quad (3.41)$$

In the actual distribution problem we have $m_j = m+j-1$ $(j = p, \ldots, 1)$ and hence (3.39) and (3.41) reduce respectively to

$$\begin{aligned}
&\mathrm{F}\{(m+p-1, n \,;\, . \,.\, m, n) \,;\, \mathrm{x}\} \\
&\quad = -\mathrm{F}\{(m+p-2, n \,;\, . \,.\, m, n) ; \mathrm{x}\} \times (1/n-m-p)\, \mathrm{F}_0(m+p-1, n-1 \,;\, \mathrm{x}) \\
&\qquad + 2 \sum_{p=s-1}^{1} (-1)^{p-s-1}\, \mathrm{F}\{(m+p-2, n \,;\, . \,.\, m+s, n \,;\, m+s-2, n \,;\, . \,.\, m, n) \,;\, \mathrm{x}\}\mathrm{Q}_s
\end{aligned}$$

where $$\mathrm{Q}_s = (1/n-m-p)\mathrm{F}(2m+p+s-2, 2n-1 \,;\, \mathrm{x}) \quad .. \quad (3.42)$$

and

$$\begin{aligned}
&\mathrm{F}\{(m+p-1, n \,;\, . \,.\, m, n) ; \mathrm{x}\} \\
&\quad = (-1)^{p-1}\, \mathrm{F}\{(m+p-1, n \,;\, . \,.\, m+1, n) \,;\, \mathrm{x}\}\ (1/m+1)\ \mathrm{F}_0(m+1, n-1 \,;\, \mathrm{x}) \\
&\qquad + (-1)^{p} \sum_{s=p}^{2} (-1)^{p-s}\, \mathrm{F}\{(m+p-1, n ; \,.\,.\, m+s, n \,;\, m+s-2, n \,;\, . \,.\, m+1, n) \,;\, \mathrm{x}\}2\mathrm{Q}_9
\end{aligned}$$

where $$\mathrm{Q}_9 = (1/m+1)\mathrm{F}(2m+s, 2n-1 \,;\, \mathrm{x}) \quad .. \quad (3.43)$$

We shall now apply (3.34) for the maximum statistic and (3.42) for the minimum statistic to the particular case of three or four variates, that is, for $p = 3$ and $p = 4$.

For $p = 3$, (3.34) becomes $\mathrm{F}\{\mathrm{x} \,;\, (m+2, n \,;\, m+1, n \,;\, m, n)\}$

$$\begin{aligned}
&= -\mathrm{F}\{\mathrm{x} \,;(m+1, n \,;\, m, n)\}\ (1/n-m-3)\ \mathrm{F}_0(\mathrm{x} \,;\, m+2, n-1) \\
&\quad + (2/n-m-3)\mathrm{F}(\mathrm{x} \,;\, 2m+3, 2n-1)\mathrm{F}(\mathrm{x} \,;\, m, n) \\
&\quad - (2/n-m-3)\ \mathrm{F}(\mathrm{x} \,;\, 2m+2, 2n-1)\mathrm{F}(\mathrm{x}, m+1, n) \\
&= (1/n-m-3)[2\mathrm{F}(\mathrm{x} \,;\, 2m+3, 2n-1)\mathrm{F}(\mathrm{x} \,;\, m, n) - 2\mathrm{F}(\mathrm{x} \,;\, 2m+2, 2n-1)\mathrm{F}(\mathrm{x} \,;\, m+1, n) \\
&\quad - (1/n-m-2)\ \mathrm{F}_0(\mathrm{x} \,;\, m+2, n-1)(-\mathrm{F}_0(\mathrm{x} \,;\, m+1, n-1)\ \mathrm{F}(\mathrm{x} \,;\, m, n) \\
&\quad + 2\mathrm{F}(\mathrm{x} \,;\, 2m+1, 2n-1)] \quad .. \quad (3.44)
\end{aligned}$$

and (3.42) becomes $\mathrm{F}\{(m+2, n \,;\, m+1, n \,;\, m, n) \,;\, \mathrm{x}\}$

$$\begin{aligned}
&= -(1/n-m-3)\ \mathrm{F}_0(m+2, n-1 \,;\, \mathrm{x})\ \mathrm{F}\{(m+1, n \,;\, m, n) \,;\, \mathrm{x}\} \\
&\quad + (2/n-m-3)\ [\mathrm{F}(2m+3, 2n-1 \,;\, \mathrm{x})\ \mathrm{F}(m, n \,;\, \mathrm{x}) \\
&\quad - \mathrm{F}(2m+2, 2n-1 \,;\, \mathrm{x})\mathrm{F}(m+1, n \,;\, \mathrm{x})] \\
&= -(1/n-m-3)[(1/n-m-2)\ \mathrm{F}_2(m+2, n-1 \,;\, \mathrm{x})(-\mathrm{F}_0(m+1, n-1 \,;\, \mathrm{x}) \\
&\qquad \times \mathrm{F}(m+1, n \,;\, \mathrm{x}) + 2\mathrm{F}(2m+2, 2n-1 \,;\, \mathrm{x})) \\
&\quad + \mathrm{F}(2m+3, 2n-1 \,;\, \mathrm{x})\mathrm{F}(m, n \,;\, \mathrm{x}) - \mathrm{F}(2m+2, 2n-1 \,;\, \mathrm{x})\mathrm{F}(m+1, n \,;\, \mathrm{x})] \quad .. \quad (3.45)
\end{aligned}$$

For $p = 4$, (3.34) becomes $\mathrm{F}\{\mathrm{x} \,;\, (m+3, n \,;\, m+2, n \,;\, m+1, n \,;\, m, n)\}$

$$\begin{aligned}
&= (1/n-m-4)[-\mathrm{F}_0(\mathrm{x} \,;\, m+3, n-1 \times (3.44) \\
&\quad + (2/n-n-3)\ \mathrm{F}(\mathrm{x} \,;\, 2m+5, 2n-1 ; \mathrm{x})\ (-\mathrm{F}_0(\mathrm{x} \,;\, m+1, n-1)\mathrm{F}(\mathrm{x} ; m, n) \\
&\quad + 2\mathrm{F}(\mathrm{x} \,;\, 2m+1, 2n-1)) \\
&\quad + (2/n-m-3)\mathrm{F}(\mathrm{x} \,;\, 2m+3, 2n-1)(-\mathrm{F}_0(\mathrm{x} \,;\, m+2, n-1)\mathrm{F}(\mathrm{x} \,;\, m+1, n) \\
&\quad + 2\mathrm{F}(\mathrm{x} \,;\, 2m+3, 2n-1)) \\
&\quad - (2/n-m-3)\mathrm{F}(\mathrm{x} \,;\, 2m+4, 2n-1)(-\mathrm{F}_0(\mathrm{x} \,;\, m+2, 2n-1)\ \mathrm{F}(\mathrm{x} \,;\, m, n) \\
&\quad - (m+2/n-m-2)\ \mathrm{F}_0(\mathrm{x} \,;\, m+1, n-1)\mathrm{F}(\mathrm{x} ; m, n) \\
&\quad + 2\mathrm{F}(\mathrm{x} \,;\, 2m+3, 2n-1) + (2m+4/n-m-2)\mathrm{F}(\mathrm{x} \,;\, 2m+2, n-1) \quad .. \quad (3.46)
\end{aligned}$$

For $p = 4$, (3.42) becomes $\mathrm{F}\{(m+3, m ; m+2, n ; m+1, n ; m,n) ; \mathrm{x}\}$

$$\begin{aligned}
&= -(1/n-m-4)[\mathrm{F}_0(m+3, n-1 ; \mathrm{x}) \times (3.45) \\
&\quad +(2/n-m-3)\mathrm{F}(2m+5, n-1 ; \mathrm{x})(-\mathrm{F}_0(m+1, n-1 ; \mathrm{x})\, \mathrm{F}(m,n ; \mathrm{x}) \\
&\quad +2\mathrm{F}(2m+1, 2n-1 ; \mathrm{x}))\ (+(2/n-m-3)\mathrm{F}(2m+3, 2n-1 ; \mathrm{x}) \\
&\quad \times(-\mathrm{F}_0(m+2, n-1 ; \mathrm{x})\, \mathrm{F}(m+1, n ; \mathrm{x})+2\mathrm{F}(2m+3, 2n-1 ; \mathrm{x})) \\
&\quad -(2/n-m-3)\mathrm{F}(2m+4 ; 2n-1 ; \mathrm{x})(-\mathrm{F}_0(m+2, n-1 ; \mathrm{x})\mathrm{F}(m,n ; \mathrm{x}) \\
&\quad +(m+2/n-m-2)\mathrm{F}_0(m+1, n-1 ; \mathrm{x})\, \mathrm{F}(m,n ; \mathrm{x}) \\
&\quad +2\mathrm{F}(2m+2, 2n-1 ; \mathrm{x})+2(m+2/n-m-2)\mathrm{F}(2m+1, 2n-1 ; \mathrm{x}))] \qquad .. \quad (3.47)
\end{aligned}$$

We can if we like also apply (3.42) and (3.43) to find out the actual forms for the cases $p=3$ and $p = 4$ when reduction up to the final stage (that is, where we can directly use the incomplete B-fuction tables) has been effected. This, however, would appear to be hardly necessary for purposes of illustration after the detailed reduction already given by (3.46) and (3.47). It is more worth while noting from (3.34), (3.35), (3.42), (3.43) that for final reduction of the p-variate problem it is only necessary to calculate

$$\begin{aligned}
&\mathrm{F}(\mathrm{x} ; m+j, n-1) \equiv \mathrm{F}(m+j ; n-1 ; \mathrm{x})(j=p, p-1, .. 1),\ \mathrm{F}(\mathrm{x} ; 2m+i, 2n-1),\ \mathrm{F}(2m+i, 2n-1; \mathrm{x}) \\
&\qquad (i=2p-2, 2p-3, \ldots 1),\ \mathrm{F}(\mathrm{x} ; m+j' ; n-1),\ \mathrm{F}(m+j', n-1 ; \mathrm{x})(j'=p-1, \ldots 1)
\end{aligned}$$

It is evident from a look at (3.34), (3.35), (3.42) and (3.43) as well as at the more general forms (3.3), (3.33), (3.39) and (3.42) that once the basic quantities indicated just now have been calculated from the incomplete B-function tables, all functions like

$$\begin{aligned}
&\mathrm{F}\{(m+p-1, n; m+p-2, n; \ldots m,n); \mathrm{x}\},\ \mathrm{F}\{(m+p-2, n; \ldots m+p-3, n; \ldots m,n); \mathrm{x}\}, \\
&\qquad \ldots.\ \mathrm{F}\{(m+1, n ; m,n) ; \mathrm{x}\} \text{ and also } \mathrm{F}\{\mathrm{x} ; (m+p-1, n ; \ldots m,n)\}, \\
&\qquad \ldots.\ \mathrm{F}\{\mathrm{x} ; (m+p-2, n, \ldots m,n)\},\ \mathrm{F}\{\mathrm{x} ; (m+1, n ; m,n)\}
\end{aligned}$$

can be evaluated. A slight adaptation and adjustment of the recursion formalae to the needs of a compact computational procedure would be, however, necessary. This is possible and has been actually done, but is not worth while indicating in the present paper. It will, however, be discussed in detail and the actual tabulational and computational procedure will be set forth in a later paper by the author and other collaborators.

4. THE DERIVATION OF THE DISTRIBUTION OF $x_j(j = p-1, p-2, \ldots, 2)$

We shall consider now the functions.

$$\mathrm{F}\{m_p,n ; m_{p-1},n ; \ldots m_{j+1},n ; m_j,n ; \mathrm{x} ; m_{j-1},n ; \ldots m_1,n\} \ldots (j=p-1, p-2, \ldots, 2)$$

$$\equiv \mathrm{F}\left\{\begin{array}{lllll}
 & m_p,n & m_{p-1},n & .. & m_1,n \\
 & . & . & .. & . \\
 & m_p,n & m_{p-1},n & .. & m_1,n \\
\mathrm{x}; & (m_p,n & m_{p-1},n & .. & m_1,n) \\
 & m_p,n & m_{p-1},n & .. & m_1,n \\
 & . & . & .. & . \\
 & m_p,n & m_{p-1},n & .. & m_1,n
\end{array}\right\} \qquad .. \quad (4)$$

Suppose that we tried to reduce m_p successively by 1 ; the pseudo-determinant is easily expanded into the sum of a number of terms of which the typical one is either of the following different types.

(*i*) $\mathrm{F}(m'_p, n ; \ldots m'_j, n ; \mathrm{x} ; m'_{j-1}, n ; \ldots m'_s, n ; \ldots m'_1, n)$ ..

(*ii*) $\mathrm{F}(m'_p, n ; \ldots m'_s, n ; \ldots m'_j, n ; \mathrm{x} ; m'_{j-1}, n \ldots m'_1, n)$ ....

(*iii*) $\mathrm{F}(m'_p, n ; \ldots m'_s, n ; \mathrm{x} ; m'_{j-1}, n ; \ldots m'_1,n)$

where $m'_s = m_p$ and where in the three different cases we have respectively,

$$(i)\ s<j \qquad (ii)\ s>j \qquad (iii)\ s=j$$

The whole pseudo-deteminant is $\Sigma(i)$, $\Sigma(ii)$ or $\Sigma(iii)$ where the summation is taken over all permutations of $(m_p, m_{p-1}, \dots m_1)$, a typical permutation being denoted as before by $(m'_p, m'_{p-1} \dots m'_2, m'_1)$. It follows from (2.81), (2.82) and (2.84) that

$$\begin{aligned}
&(i)\ \ F(m'_p,n\,;\,..\,m'_j,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_s,n\,;\,..\,m'_1,n)\\
&=-(1/n-m_p-1)\ F(m_p,n\,;\,..\,m'_j,n\,;\,x\,;\,m'_{j-1},n;\,..\,m'_{s+1}+m_p,\,2n-1\,;\,m'_{s-1},n\,;\,..\,m'_1,n)\\
&\quad+(1/n-m_p-1)\ F(m_p,n\,;\,..\,m'_j,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_{s+1},n\,;\,m'_{s-1}+m_p,2n-1\,;\,..\,m'_1,n)\\
&\quad+(m_p/n-m_p-1)\ F(m_p,n\,;\,..\,m'_j,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_{s+1},n\,;\,m_p-1,n\,;\,..\,m'_1,n)\quad ..\quad (4.1)
\end{aligned}$$

except when $s = 1$ in which case the second term would be absent. In the notation previously introduced we can write this in the alternative form.

$$\begin{aligned}
&-(1/n-m_p-1)\,F(m'_p,n;\,..\,m'_j,n\,;\,x\,;\,m'_{j-1},n;\,..\,m'_{s+1},n\,;\,\overleftarrow{m}_p,\,\overleftarrow{n}-1\,;\,m'_{s-1},n;\,..\,m'_1,n)\\
&+(1/n-m_p-1)\ F(m'_p,n;\,..\,m'_j,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_{s+1},n\,;\,\overrightarrow{m}_p,\overrightarrow{n}-1\,;\,m'_{s-1},n\,;\,..\,m'_1,n)\\
&+(m_p/n-m_p-1)F(m'_p,n;\,..\,m'_j,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_{s+1},n\,;\,m_p-1,\,n;\,..\,m'_1,n)\quad ..\quad (4.11)
\end{aligned}$$

Similarly

$$\begin{aligned}
&(ii)\ \ F(m'_p,n\,;\,..\,m'_{s+1},n\,;\,m_p,n\,;\,m'_{s-1},n\,;\,..\,m'_j,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_1,n)\\
&=\ (1/n-m_p-1)\,F(m'_p,n\,;\,..\,m'_{s+1},n\,;\,\overrightarrow{m}_p,\overrightarrow{n}-1\,;\,m'_{s-1},\,n\,;\,..\,m'_j,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_1,n)\\
&\quad-(1/n-m_p-1)\,F(m'_p,n\,;\,..\,m'_{s+1},n\,;\,,\overleftarrow{m}_p,\overleftarrow{n}-1\,;\,m'_{s-1},n\,;\,..\,m'_jn\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_1,n)\\
&\quad+(m_p/n-m_p-1)\,F(m'_p,n;\,..\,m'_{s+1},n;\,m_p-1,n;\,m'_{s-1},n;\,..\,m'_j,n;\,x\,;\,m'_{j-1},n;\,..\,m'_1,n)\quad (4{\cdot}12)
\end{aligned}$$

except when $s = p$ in which case the second term would be absent, and finally

$$\begin{aligned}
&(iii)\ \ F(m'_p,n\,;\,..\,m'_{j+1},n\,;\,m_p,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_1,n)\\
&\qquad=\ (1/n-m_p-1)\ F_0(x\,;\,m_p,n-1)\ F(m'_p,n\,;\,..\,m'_{j+1},n\,;\,x)\ F(x\,;\,m'_{j-1},n\,;\,..\,m'_1,n)\\
&\qquad\quad+(1/n-m_p-1)\ F(m'_p,n\,;\,..\,m'_{j+1},n\,;\,m'_{j-1}+m_p,2n-1\,;\,x\,;\,..\,m'_1,n)\\
&\qquad\quad-(1/n-m_p-1)\ F(m'_p,n\,;\,..\,m'_{j+1}+m_p,2n-1\,;\,x\,;\,m'_{j-1}n\,;\,..\,m'_1,n)\\
&\qquad\quad+(m_p/n-m_p-1)\ F(m'_p,n\,;\,..\,m'_{j+1},n\,;\,m_p-1,\,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_1,n)
\end{aligned}$$

or in a different notation

$$\begin{aligned}
&=\ (1/n-m_p-1)\ F_0(x\,;\,m_p,n-1)\ F(m'_p,n\,;\,..\,m'_{j+1},n\,;\,x)\ F(x\,;\,m'_{j-1},n\,;\,..\,m'_1,n)\\
&\quad+(1/n-m_p-1)\ F(m'_p,n\,;\,..\,m'_{j+1},n\,;\,m_p,n-1\,;\,x\,;\,\overleftarrow{m}'_{j-1},\overleftarrow{n}\,;\,..\,m'_1,n)\\
&\quad-(1/n-m_p-1)\ F(m'_p,n\,;\,..\,m'_{j+1},n\,;\,\overleftarrow{m}_p,\overleftarrow{n}-1\,;\,x\,;\,m'_{j-1},n-1\,;\,..\,m'_1,n)\\
&\quad+(m_p/n-m_p-1)\ F(m'_p,n\,;\,..\,m'_{j+1},n\,;\,m_p-1,n\,;\,x\,;\,m'_{j-1},n\,;\,..\,m'_1,n)\quad ..\quad (4.13)
\end{aligned}$$

The mechanism by which (4.13) is obtained becomes evident if we consider (for purposes of illustration) the case of three variables

$$\int_{x_2}^{\infty}\frac{x_3^{m'_3}\,dx_3}{(1+x_3)^n}\int_{x}^{\infty}\frac{x_2^{m_3}\,dx_2}{(1+x_2)^n}\int_{0}^{x_2}\frac{x_1^{m'_1}\,dx_1}{(1+x_1)^n}=\int_{x}^{\infty}\frac{x_2^{m_3}\,dx_2}{(1+x_2)^n}\cdot\phi(x_2)\,\psi(x_2),$$

where $$\phi(x_2)=\int_{x_2}^{\infty}\frac{x_3^{m'_3}\,dx_3}{(1+x_3)^n}\ \text{and}\ \psi(x_2)=\int_{0}^{x_2}\frac{x_1^{m'_1}\,dx_1}{(1+x_1)^n}\quad ..\qquad (4.14)$$

Then integrating by parts (the integration being for $1/(1+x_2)^{n-m_3}$ and the differentiation for $\varphi(x_2)\,\psi(x_2)..x_2^{m_3}/(1+x_2)^{m_3}$ we should have the expression

$$=(1/n-m_3-1)\{x^{m_3}/(1+x)^{n-1}\}\,\phi(x)\,\psi(x)+(1/n-m_3-1)\int_x^{\infty}\frac{x_2^{m_3}\,dx_2}{(1+x_2)^{n-1}}[\varphi(x_2)\,\psi'(x_2)+\phi'(x_2)\,\psi(x_2)]$$

$$+(m_3/n-m_3-1)\int_x^{\infty}\frac{x_2^{m_3-1}\,dx_2}{(1+x_2)^{n}}\,\varphi(x_2)\,\psi(x_2)$$

$$=\ (1/n-m_3-1)\,F_0(x\,;m_3,n-1)\,F(m'_3,n\,;x)\,F(x\,;m'_1,n)$$
$$+(1/n-m_3-1)\,F(m'_3,n\,;m_3+m'_1,2n-1\,;x)$$
$$-(1/n-m_3-1)\,F(m'_3+m_3,2n-1\,;x\,;m'_1,n)+(m_3/n-m_3-1)\,F(m'_3,n\,;m_3-1,n\,;x\,;m_1',n)$$
$$=\ (1/n-m_3-1)\,F_0(x\,;m_3,n-1)\,F(m'_3,n\,;x)\,F(x\,;m'_1,n)$$
$$+(1/n-m_3-1)\,F(m'_3,n\,;m_3,n-1\,;x\,;\overleftarrow{m'_1},\overleftarrow{n})$$
$$-(1/n-m_3-1)\,F(m'_3,n\,;\overleftarrow{m_3},\overleftarrow{n}-1.\,x\,;m'_1,n)$$
$$+(m_3/n-m_3-1)\,F(m'_3,n\,;m_3-1,n\,;x\,;m'_1,n)\qquad ..\quad (4.15)$$

Hence we have

$$F\left\{\begin{array}{llll} m_p,n & m_{p-1},n & .. & m_1,n\\ m_p,n & m_{p-1},n & .. & m_1,n\\ . & . & .. & .\\ x;\ (m_p,n & m_{p-1},n & .. & m_1,n)\\ . & . & .. & .\\ m_p,n & m_{p-1},n & .. & m_1,n \end{array}\right\}=(1/n-m_p-1)\ F\left\{\begin{array}{llll} \rightarrow\ \rightarrow & & & \\ m_p,n-1 & m_{p-1},n & .. & m_1,n\\ \rightarrow\ \rightarrow & & & \\ m_p,n-1 & m_{p-1},n & .. & m_1,n\\ \rightarrow\cdot\rightarrow & . & .. & .\\ x;(m_p,n-1 & m_{p-1},n & .. & m_1,n)\\ \rightarrow\cdot\rightarrow & . & .. & .\\ m_p,n-1 & m_{p-1},n & .. & m_1,n\\ \square & m_{p-1},n & .. & m_1,n \end{array}\right\}$$

$$-(1/n-m_p-1)\ F\left\{\begin{array}{llll} \square & m_{p-1},n & .. & m_1,n\\ \leftarrow\ \leftarrow & & & \\ m_p,n-1 & m_{p-1},n & .. & m_1,n\\ \leftarrow\cdot\leftarrow & . & .. & .\\ x;(m_p,n-1 & m_{p-1},n & .. & m_1,n)\\ \leftarrow\cdot\leftarrow & . & .. & .\\ m_p,n-1 & m_{p-1},n & .. & m_1,n \end{array}\right\}+(m_p/n-m_p-1)F\left\{\begin{array}{llll} m_p-1,n & m_{p-1},n & .. & m_1,n\\ m_p-1,n & m_{p-1},n & .. & m_1,n\\ . & . & .. & .\\ x;(m_p-1,n & m_{p-1},n & .. & m_1,n)\\ . & . & .. & .\\ m_p-1,n & m_{p-1},n & .. & m_1,n \end{array}\right\}$$

$$+(1/n-m_p-1)\,F_0(x\,;m_p,n-1)[F\{(m_{p-1},n;\,m_{p-2},n;..m_j,n);\,x\}\,F\{x;\,(m_{j-1},n;\,m_{j-2},n;..m_1,n)\}$$
$$-F\{(m_{p-1},n\,;m_{p-2},n\,;\,..\,m_{j+1},n\,;m_{j-1},n)\,;x\}\ F\{x\,;(m_j,n\,;m_{j-2},n\,;\,..\,m_1,n)\}\pm..]\qquad (4.16)$$

$$=\ (2/n-m_p-1)\sum_{s=p-1}^{j}(-1)^{p-s}\,F\{m_{p-1},n;\,..\,m_{s+1},n;\,m_{s-1},n;\,..\,m_j,n;\ x;\,m_{j-1},n;\,..\,m_1,n\}R_1$$

$$+(2/n-m_p-1)\sum_{s'=j-1}^{1}-(1)^{p-1-s'}\,F\{m_{p-1},n;\,..\,m_j,n;\,x;\,m_{j-1},n;\,..\,m_{s'+1},n\,;m_{s'-1},n;\,..\,m_1,n\}R_2$$
$$+(m_p/n-m_p-1)\ F\{m_p-1,n;\,m_{p-1},n;..m_j,n;\ x;\ m_{j-1},n;..m_1,n\}$$
$$+(1/n-m_p-1)\,F_0(x;\,m_p,n-1)\ \Sigma\pm F\{(m'_{p-1},n\,;m'_{p-2},n\,;\,..\,m'_{j+1},n\ m'_j,n);x\}R_3\qquad ..\quad (4.17)$$

where $\qquad R_1=F(m_s+m_p,\,2n-1\,;x),\ R_2=F(x;\,m_{s'}+m_p,\,2n-1)$

and $\qquad R_3=F\{x\,;(m'_{j-1},n\,;m'_{j-2},n\,;\,..\,m'_1,n)\}$

and where (*i*) $m'_{p-1}>m'_{p-2}>..>m'_{j+1},>m'_j$ (*ii*) $m'_{j-1}>m'_{j-2}>..m'_1$ and (*iii*) $m'_{p-1},m'_{p-2},..m'_j$ is any selection subject to (*i*), of $p-1-(j-1)$, that is, $p-j$ out of $m_{p-1},m_{p-2}\,....\,m_1(m_{p-1}>\,...\,>m_1)$ and (*iv*) $m'_{j-1},\,...\,m'_1$ are those of $m_{p-1},\,...,\,m_1$ that are left out after the selection (*iii*). We take the +*ve* or −*ve* sign in this summation, acording as the corresponding term in the formal determinantal expansion is +*ve* or −*ve*.

By successive reduction this comes out as

$$2\sum_{s=p-1}^{j}(-1)^{p-1-s}F\{m_{p-1},n\,;\,..\,m_{s+1},n\,;\,m_{s-1},n\,;\,..\,m_j,n\,;\,x\,;\,m_{j-1},n\,;\,..\,m_1,n\}\sum_{r=1}^{m_p-m_{p-1}}R$$

$$+2\sum_{s'=j-1}^{1}(-1)^{p-1-s'}F\{m_{p-1},n\,;\,..\,m_j,n\,;\,x;\,m_{j-1},n\,;\,..\,m_{s'+1},n\,;\,m_{s'-1},n\,;\,..\,m_1,n\}\sum_{r=1}^{m_p-m_{p-1}}R_4$$

$$+\Sigma\pm F\{(m'_{p-1},n\,;\,m'_{p-2},n\,;\,..\,m'_j,n);\,x\}\;F\{x\,;\,(m'_{j-1},n\,;\,..\,m'_1n)\}\sum_{r=1}^{m_p-m_{p-1}}R_5 \quad (4\cdot18)$$

where $R = ({}_{m_p}P_{r-1}/{}_{n-m_p-2+r}P_r)\;F(m_s+m_p-r+1,\;2n-1\;;\;x),$

$R_4 = ({}_{m_p}P_r/{}_{n-m_p-2+r}P_r)\;F(x\,;\;m_{s'}+m_p-r+1,\,2n-1)$

and $R_5 = ({}_{m_p}P_{r-1}/{}_{n-m_p-2+r}P_r)\;F(x;\,m_p-r+1,n-1)$

If on the other hand we had tried to increase m_1 successively by one to m_2 then we would have had.

$$F\{m_p,n\,;\,..\,m_j,n\,;\,x\,;\,m_{j-1}n\,;\,..\,m_1,n\}$$

$$=(-1)^p\sum_{s=p}^{j}(-1)^{p-s}F\{m_p,n\,;\,..\,m_{s+1},n\,;\,m_{s-1},n\,;\,..\,m_j,n\,;\,x\,;\,m_{j-1},n\,;\,..\,m_2,n\}\sum_{r=1}^{m_2-m_1}R_6$$

$$+(-1)^p\sum_{s'=j-1}^{2}(-1)^{p-s'}F\{m_p,n\,;\,..\,m_j,n;\,x\,;\,m_{j-1},n\,;\,..\,m_{s'+1},n\,;\,m_{s'-1},n\,;\,..\,m_2,n)\sum_{r=1}^{m_2-m_1}R_7$$

$$+\Sigma\pm F\{m'_p,n\,;\,m'_{p-1},n\,;\,..\,m'_{j+1},n);\,x\}\;F\{x\,;\,(m'_j,n\,;\,...\,m'_2n)\}\sum_{r=1}^{m_2-m_1}R_8$$

where $R_6=({}_{n-m_1-r}P_{r-1}/{}_{m_1+r}P_r)\;F(m_1+r-m_s,\;2n-1;\,x),$

$R_7=({}_{n-m_1-r}P_{r-1}/{}_{m_1+r}P_r)\;F(m_1+r+m_{s'},\,2n-1\;;\;x)$

and $R_8=({}_{n-m_1-r}P_{r-1}/{}_{m_1+r}P_r)\;F_0(x\,;\,m_1+r,\,n-1)$ .. (4.19)

and where $m'_p,\;m'_{p-1},\,\ldots,\,m'_{j+1}$ is any selection (in proper order) of $(m_p,\,m_{p-1},\,\ldots,\,m_2)$ and $(m'_j\ldots m'_2)$ is the complementary set.

In actual statistical problems we have $m_i=m+i-1$ $(i=p,\;p-1,\;\ldots\;1)$ and hence $m_p-m_{p-1}=1$, and in (4.18) and (4.19) there would be only one term each in the 'r' summations although the numbers of terms in the 's' summations would be the same (for any 'r', which latter is only one here). We shall consider in particular the case of $p=3$ for which

$$\begin{aligned}F\{m+2,\,n\,;\,m+1,\,n\,;\,x\,;\,m,n\}&=(2/n-m-3)[F(m+1,\,n\,;\,x)F(2m+2,\,2n-1\,;\,x)\\&\quad-F(x\,;\,m,\,n)F(x\,;\,2m+3,\,n)]+(1/n-m-3)F_0(x\,;\,m+p-2,\,n-1)\\&\quad\times[(F(m+1,\,n\,;\,x\,)F(x\,;\,m,\,n)-F(m,\,n\,;\,x)\,F(x\,;\,m+1,\,n)] \quad ..\;(4.2)\\ \text{or}\quad&=-(2/m+1)\,[F(m+2,\,n\,;\,x)\,F(2m+1,\,2n-1\,;\,x)-F(x\,;\,m+1,\,n)\,F(x\,;\,2m+2,2n-1)]\\&\quad+(1/m+1)\,F_0(x\,;\,m+1,\,n-1)\,[F(m+2,\,n\,;\,x)\,F(m+1,\,n\,;\,x)-F(\;\;)F(\;\;)]\;..\quad(4.21)\end{aligned}$$

REPERENCES

1. Fisher, R. A. (1939) The sampling distributions of some statistics obtained from non-linear equations. *Annals of Eugenics*, **3**, 238-244.

2. Hsu, P. L. (1939) On the distributions of roots of certain determinantal equations. *Annals of Eugenics*, **3**, 250-258.

3. Roy, S. N. (1939) P-statistics or some generalisations in analysis of variance appropriate to multivariate problems. *Sankhyā*, **3**, 341-396.

4. Roy, S. N. (1942) The sampling distribution of p-statistics and certain allied statistics on the non-null hypothesis. *Sankhyā*, **1**, 15-34.

5. Roy, S. N. (1942a) Analysis of variance for multivariate normal populations. The sampling distribution of the requisite p-statistics on the null and non-null hypotheses. *Sankhyā*, **6**, 35-50.

Paper received: 30 November, 1943.

DISTRIBUTIONS OF CERTAIN FREQUENCY CONSTANTS IN SAMPLES FROM NON-NORMAL POPULATIONS*

By K. C. CHERIYAN
Statistical Laboratory, Calcutta

INTRODUCTION

In this paper a study of the sampling distribution of the correlation coefficient in samples from correlated Pearsonian type III populations is made by means of experimental sampling. Using the transformation $z=\frac{1}{2}\{\log_e (1+r)-\log_e (1-r)\}$ to the observed values of r, the sampling distributions of z are obtained. The frequencies and frequency constants of these distributions are then compared with corresponding 'normal theory' values. The sampling distributions of variance and mean are also recorded in order to see how far they agree with known theoretical results.

THE DISTRIBUTION OF THE CORRELATION COEFFICIENT

The historical background to this problem may be briefly described as follows. The theoretical distribution of the correlation coefficient, when samples are drawn from an infinite bivariate normal population, was given by Fisher in 1915. As the mathematical expression defining this distribution was not in a form suitable for the evaluation of frequencies in different class-intervals, 'A Co-operative Study' was made by several authors (1915) to construct tables giving the relative frequencies for different values of r. Later in 1929, E. S. Pearson showed that Fisher's z-transformation could be used to obtain frequencies in the different class-intervals of the r-distribution by entering values of $(z-\text{mean } z)/\sigma_z$ in any table of the normal probability integral. In particular, he took the first and second approximations for the values of mean z and σ^2_z. These two methods have since been used to compare values of r obtained in samples from non-normal populations with normal theory values. Student's work (1908) in this direction is prior to that of Fisher, but the population he used for experimental sampling was only approximately normal. E.S. Pearson (1929 & 1932) has contributed several articles in which it is observed that for parent populations with different values of ρ, there is close agreement between observation and 'normal theory'. The same result was obtained by Rider (1932) who used rectangular and triangular populations. But as Hey (1928) points out, "No attempt appears to have been made to carry out experimental sampling from a bivariate population in which the distribution surface is not normal and in which the correlation is high". Since the theoretical distribution of r has been established only in the case where the parent population is normal, experimental sampling from all other types of populations will be welcome as they help to determine whether tests of significance based on 'normal theory' assumptions are relevant or not.

EXPERIMENTAL WORK

In this paper, bivariate type III populations with varying degrees of skewness and correlation are constructed from the squares of normal deviates.† Since each of the squared values has a distribution given by the law, $y= y_0\, x^{-1/2}\, e^{-x}$, the sum of n of these will be distributed according to the law, $y = y_0 x^{(n-1)/2}\, e^{-x}$. Thus, if it be required to construct a population in two variables with $\rho = 3/4$, we take groups of five from among the squares of normal deviates, add the first four and the last four and denote them x and y. This device provides an easy way of constructing populations with any degree of skewness and any desired value for ρ. In another paper (1941) I have obtained mathematical expressions for this type of frequency distributions as also for their moment generating functions. In a recent paper dealing with the same problem Krishnan Nair (1941) has constructed two populations with marginal distributions $y = y_0 e^{-x}$ and $y = y_0 . xe^{-x}$. In each case $\rho = 0$. It is observed that in both cases normal theory values provide good fit to sampling distributions of r, and also much better fit than either of the approximations. The problem attempted here is a continuation of this work by sampling from populations in which the variates are highly correlated.

---

*Part of a thesis approved by the University of Madras for the M. Sc. Degree.

†This table was formed by A. N. Krishnan Nair (1941).

The following three bivariate populations were constructed according to the method outlined above. The betas and skewnesses in the following table are those of the marginal distributions.

	Marginal Distribution	β_1	β_2	Skewness	Correlation
Population I	$y=y_0\ e^{-x}$	4·0	9·0	1·0000	1/2
Population II	$y=y_0\ x\ e^{-x}$	2·0	6·0	0·7070	3/4
Population III	$y=y_0\ x^{7/2}\ e^{-x}$	0·9	4·3	0·4814	8/9

Two hundred samples of five were taken from each of the three populations and the correlation coefficient was calculated for each sample. Thus we get three experimental distributions of r, the total frequency in each distribution being 200. To find how far these distributions show agreement with 'normal theory', we obtain the frequencies for the latter from the tables provided in 'A co-operative study' (13). Frequencies were also obtained by employing the following approximations for the means and variances of transformed r, and using Sheppard's tables of the normal probability integral.

Approximation I.

$$\text{Mean } z = \xi + \frac{\rho}{2(n-1)}, \quad \sigma^2_z = \frac{1}{n-3}$$

Approximation II.

$$\text{Mean } z = \xi + \frac{\rho}{2(n-1)} \left\{1 + \frac{1+\rho^2}{8(n-1)} + \ldots\right\}$$

$$\sigma^2_z = \frac{1}{n-1} \left\{1 + \frac{4-\rho^2}{2(n-1)} + \frac{176-21\rho^2-21\rho^4}{48(n-1)^2} + \ldots \right\}$$

Table 1 gives the theoretical and observed distributions of r. Table 2 gives the observed and expected values of the first two moments and betas.

It is seen that in these tables, the χ^2-values indicate very poor agreement between observation and normal theory. But examining the tables more closely, we find that this is mainly due to a few unusually high values among the observed frequencies. This might very well be due to errors of sampling or an inadequate total frequency. In all other class-intervals, the agreement between observation and theory is remarkably good. Further, as we are sampling from populations with high positive correlations, we are only concerned with the distribution of the observed frequencies in the immediate neighbourhood of the theoretical value. Considering the total frequencies at either end of the range, we have the grouped values shown in Table 3. These results are in fairly close agreement between observation and theory, but it should be noted that as the correlation increases, the agreement becomes less satisfactory.

TABLE 1. DISTRIBUTION OF r FOR POPULATIONS I, II AND III

central values of r	Population I ($\rho=0\cdot5$)				Population II ($\rho=0\cdot75$)				Population III ($\rho=8/9$)			
	observed frequencies	expected			observed frequencies	expected			observed frequencies	expected		
		normal theory	approximation			normal theory	approximation			normal theory	approximation	
			I	II			I	II			I	II
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
−0·95	1	0·5	0·3	0·1	1	0·1	0·1	0·2	—	0·0	0·0	0·-
−0·85	—	1·1	1·2	0·7	—	0·2	0·3	0·7	—	0·0	0·0	0·1
−0·75	1	1·5	2·1	1·4	—	0·4	0·6	1·1	1	0·1	0·1	0·2
−0·65	3	2·0	2·9	2·2	1	0·5	0·6	1·4	—	0·1	0·1	0·3
−0·55	3	2·5	3·6	2·9	—	0·7	1·6	1·8	—	0·1	0·2	0·5
−0·45	1	3·1	4·3	3·7	—	0·8	1·7	2·1	—	0·2	0·2	0·6
−0·35	6	3·8	4·9	4·5	2	1·0	2·1	2·5	—	0·2	0·4	0·8
−0·25	5	4·6	5·7	5·4	—	1·3	2·7	2·9	—	0·3	0·5	1·0
−0·15	7	5·5	6·5	6·3	2	1·7	3·1	3·4	—	0·3	0·7	1·3
−0·05	7	6·6	7·3	7·3	4	2·3	4·2	3·9	—	0·6	0·9	1·6
+0·05	18	6·8	8·3	8·4	6	3·0	4·8	4·6	1	0·8	1·3	1·9
0·15	3	9·3	9·3	9·8	3	3·4	5·9	5·3	—	1·2	1·7	2·5
0·25	11	11·1	10·6	11·3	6	5·1	7·2	6·3	—	1·7	2·4	3·1
0·35	11	13·7	12·1	13·0	9	6·7	8·9	7·6	3	2·4	3·3	3·8
0·45	11	15·8	13·9	15·1	13	9·6	11·2	9·4	5	3·7	4·8	5·7
0·55	16	18·6	16·2	17·6	12	13·5	14·3	11·9	3	5·9	7·1	7·6
0·65	17	21·8	19·1	20·6	19	19·5	15·3	15·7	11	11·7	11·2	11·1
0·75	14	24·8	22·7	24·1	26	29·2	29·6	21·9	22	20·3	19·4	12·8
0·85	26	26·4	26·7	26·9	29	45·4	37·8	34·3	27	43·8	39·4	34·2
0·95	39	19·8	22·4	18·7	68	52·2	48·0	63·1	127	108·2	106·2	105·9
Total	200	199·3	200·1	200·0	200	196·6	200·0	200·1	200	201·6	199·9	200·0
χ^2		50·02	22·09	44·50		17·21	24·54	11·27		14·24	19·88	25·56
n'		15	15	15		11	11	11		8	8	8
P		0·000	0·077	0·000		0·072	0·006	0·379		0·040	0·006	0·001

TABLE 2. FREQUENCY CONSTANTS WITH CORRESPONDING 'NORMAL THEORY' VALUES

frequency constants	Population I		Population II		Population III	
	observed	expected	observed	expected	observed	expected
Mean	0·4590	0·4517	0·6770	0·7012	0·8580	0·8560
σ	0·4528	0·4239	0·3331	0·3025	0·1547	0·1927
β_1	0·6323	1·0315	3·0027	4·1720	18·3845	12·1821
β_2	2·7574	3·4191	6·6138	7·9582	31·8453	20·4275
Skewness	2·3048	1·7097	1·3761	1·9502	1·8713	2·0396

TABLE 3. GROUPED DISTRIBUTION OF r AT EITHER END OF RANGE

Total frequency		observed	normal theory	approximation	
				I	II
Population I	Below zero	34	31·4	38·8	34·5
	Above 0·7	79	71·0	71·8	69·7
,, II	Below zero	10	9·0	17·0	20·0
	Above 0·8	97	97·6	85·8	97·4
,, III	Below zero	1	1·9	3·1	6·4
	Above 0·8	154	152·0	145·6	140·1

In table 1, large deviations from expected values are found only in the class-intervals 0·0 to 0·1, 0·1 to 0·2 and 0·9 to 1·0 for Population I. For values of r between 0·2 and 0·7, the deviations from theory are very small. In the same table, for Populations II and III serious discrepancies are found to occur only towards the positive end of the range.

An important result revealed by this study is that 'normal theory' gives a fairly good fit in the immediate neighbourhood of the population correlation only when this correlation is not very high. For Population I ($\rho = 0·5$,) for a deviation from the theoretical as great as 0·25, we are justified in using normal theory for tests of significance.

For Population II ($\rho = 0·75$), a sufficiently close agreement between observation and 'normal theory' is noticeable only for values of r below 0·75, and above 0·10. For any observed correlation falling outside this range, 'normal theory' tables cannot be used, as the total frequency in these tables is not large enough for a discussion of the frequencies at the 5 and 1 per cent levels. For Population III ($\rho = 0·88$), 'normal theory' values show agreement only between 0·3 and 0·8. The observed agreement is not sufficient to justify the use of 'normal theory' assumptions.

The constants of the sampling distributions indicate very good agreement for means and standard deviations. The values of the betas are also fairly close, except for Population III. According to normal theory, the skewness of the r-distribution increases as ρ increases. But the observed skewness is maximum for the first distribution and minimum for the second.

DISTRIBUTION OF z.

The theoretical distribution of z has been shown to be approximately normal in the case where the parent population is normal. But if this property is found to persist when the parent universe is a skew one, then we can safely apply tests of significance to values of transformed r. Table 4 gives distributions of z where the z-values were obtained by entering Fisher's table (4) with the corresponding value of r.

Table 4. Distribution of z

Central values of z	Population I	Population II	Population III		Central values of z	Population I	Population II	Population III	
			observed	expected				observed	expected
Below −1·40	—	1	1	—	Below 1·70	10	12	21	24·0
−1·30	1	—	—	—	1·90	1	11	23	21·1
−1·10	1	—	—	—	2·10	—	9	19	16·9
−0·90	2	—	—	—	2·30	—	4	9	12·2
−0·70	2	1	—	0·1	2·50	—	5	9	7·9
−0·50	4	1	—	0·2	2·70	—	8	2	4·9
−0·30	9	2	—	0·4	2·90	—	2	5	2·5
−0·10	12	6	—	1·0	3·10	—	4	3	1·3
0·10	24	8	1	2·0	3·30	—	1	3	0·6
0·30	20	13	3	3·9	3·50	—	—	3	0·2
0·50	20	19	6	6·8	3·70	—	—	—	0·1
0·70	23	16	7	10·7	3·90	—	—	1	—
0·90	27	31	20	15·3	4·10	—	—	3	—
1·10	20	14	13	19·8	4·30	—	—	3	—
1·30	12	19	19	23·3	4·70	—	—	1	—
1·50	12	13	25	25·0	Total	200	200	200	200·2

$\chi^2 = 30·46$, $n' = 13$, $P = 0·001$

It is seen that the betas of the first two distributions agree well with theory. For the third distribution we observe deviations for the variance and betas. The expected

TABLE 5. PARAMETERS OF THE DISTRIBUTION OF z.

Parameter	Population I		Population II		Population III	
	observed	expected	observed	expected	observed	expected
Mean	0·6190	0·6143	1·1690	1·0713	1·7310	1·5337
μ_2	0·3920	0·4213	0·6925	0·4100	0·7595	0·4066
μ_3	−0·1616	−0·0024	0·1652	0·0000	0·4727	0·0032
μ_4	0·4530	0·5660	1·5191	0·5374	2·7799	0·5178
β_1	0·4333	0·0003	0·0655	0·0000	0·5100	0·0006
β_2	2·9479	3·7945	3·1861	3·8177	4·8192	3·6325
Skewness	0·6234	0·0059	0·1211	0·0000	0·2320	0·0089

values in this case are those of the corresponding z-distributions assuming the parent universe to be normal. They were obtained from normal probability tables by entering values of $(z-\text{ mean } z)/\sigma_z$ where the means and the standard deviations are the theoretical values. We observe that agreement is good everywhere except for values of z above 2·60. The total observed frequencies on either side of the theoretical mean are fairly equally balanced for the three distributions as is shown below

Frequency	Population I	Population II	Population III
Below mean	95	98	95
Above mean	105	102	105

In the last case we also find that agreement is good in the immediate neighbourhood of the theoretical z-values. For correlations obtained from highly correlated skew populations, the z-transformation might provide a better test of significance than the corresponding 'normal theory' tables when the observed z-value does not deviate greatly from the theoretical.

DISTRIBUTION OF VARIANCE

The exact distribution of variance in samples from non-normal populations is not known. But as Le Roux (1931) has observed, it may be supposed to conform to either type III, type V or type VI of Pearson's curves. Tchouproff (1916) first gave the formulae for moments of the variance distribution and later Church (1925) derived them again following 'Student's' shorter and more direct method. For samples of size 5, the values of the first two moments and betas are as given below.

$$M'_1 = \mu_2 \frac{4}{5} \cdot M_2 = \mu_2 \left(\frac{4}{5}\right)^2 (\beta_2 - \tfrac{1}{2})$$

$$B_1 = \frac{(\beta_4 - \frac{25\beta_2}{4} - \frac{5}{2})^2}{5(\beta_2 - \frac{1}{2})^3}$$

$$B_2 = \frac{(\beta_6 + 2\beta_4 - 26\beta_3 + \frac{\beta_2^2}{8} - \frac{195\beta_2}{4} + \frac{145\beta_1}{2} + \frac{429}{8}}{5(\beta_2 - \frac{1}{2})^2}$$

Where M'_1, M_2, B_1, B_2, are the moments and betas of the variance-distribution, and μ_2, β_1, β_2, …, β_6 are parameters of the parent population. The observed distributions of σ^2_x and σ^2_y from each of the three populations as well as the observed and expected values of constants of the variance-distribution are given in the following tables 6 and 7.

TABLE 6. FREQUENCY DISTRIBUTION OF VARIANCES OF x AND y.

Central Values	Population I		Central Values	Population II		Central Values	Population III	
	σ_x^2	σ_y^2		σ_x^2	σ_y^2		σ_x^2	σ_y^2
0·25	24	26	0·5	18	18	2·0	16	12
0·75	32	23	1·5	20	22	4·0	18	23
1·25	31	34	2·5	25	28	6·0	28	24
1·75	12	14	3·5	25	29	8·0	25	21
2·25	15	26	4·5	24	20	10·0	16	22
2·75	17	13	5·5	16	8	12·0	17	17
3·25	8	15	6·5	16	14	14·0	12	15
3·75	11	8	7·5	10	14	16·0	13	11
4·25	8	6	8·5	6	11	18·0	10	7
4·75	2	1	9·5	11	8	20·0	11	12
5·25	5	6	10·5	4	4	22·0	6	6
5·75	4	4	11·5	4	2	24·0	3	5
6·25	6	1	12·5	2	4	26·0	7	2
6·75	3	2	13·5	4	2	28·0	3	5
7·25	5	3	14·5	4	5	30·0	3	2
7·75	1	1	15·5	—	3	32·0	4	4
8·25	3	—	16·5	1	—	34·0	4	1
8·75	5	3	17·5	2	3	36·0	—	—
9·25	1	1	18·5	—	4	38·0	1	2
9·75	1	4	19·5	—	—	40·0	—	2
10·25	—	—	20·5	1	2	42·0	1	—
10·75	1	1	21·5	1	—	44·0	—	1
11·25	—	—	22·5	1	—	46·0	1	1
11·75	2	—	23·5	—	—	above 47·0	4	5
12·25	1	1	24·5	1	—			
12·75	1	2	25·5	—	1			
above 13	1	5	above 26	4	3			
Total	200	200	Total	200	200	Total	200	200

From table 7 we note that the observed B_1 and B_2 for Population I is considerably less than what might be expected. This might possibly be due to the excessive skewness in Population I. Another point of difference is that the observed standard deviations in the third case are both less than half the expected value. All the other values agree fairly well.

TABLE 7. CONSTANTS IN THE DISTRIBUTION OF THE VARIANCE.

Constants	Population I			Population II			Population III		
	observed		expected	observed		expected	observed		expected
	x	y		x	y		x	y	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mean	2·9650	2·4000	3·2000	7·1300	6·3250	6·4000	11·1450	10·9400	14·4000
$\sqrt{M_2}$	2·8437	4·1576	4·1721	6·0844	6·1409	6·7124	6·1130	5·5133	12·6085
B_1	2·3240	1·3658	18·3696	7·9944	12·0818	10·8489	6·7241	3·4766	6·4918
B_2	5·0631	4·6468	35·9585	13·6248	22·3593	20·1694	12·4583	7·2799	10·4278

DISTRIBUTION OF MEAN

The theoretical distribution of the mean when samples of size n are drawn from a type III population of the form $y = \frac{x^{p-1}e^{-x}}{\Gamma(p)}$ was proved by Irwin to be (1927) $y = \frac{ne^{-nx}\,(n\bar{x})^{p-1}}{\Gamma(np)}$ Since the marginal distributions of the parent populations used in this paper are $y = y_0\ e^{-x}$, $y = y_0\ x\ e^{-x}$ and $y = y_0\ x^{7/2}\ e^{-x}$, the corresponding mean-distributions will be given by

$$y = \{(5u)^4\ e^{-5u}\ d(5u)\}/\Gamma(5) \quad .. \quad (1)$$

$$y = \{(5u)^9\ e^{-5u}\ d(5u)\}/\Gamma(10) \quad .. \quad (2)$$

$$y = \{(5u)^{21\cdot5}\ e^{-5u}\ d(5u)\}/\Gamma(22\cdot5) \quad .. \quad (3)$$

In calculating the expected moments from these equations it has to be remembered that the unit used in the parent distribution is really $\chi^2/2$ and so for comparison with observed values, the moment μ_r obtained from this should be multiplied by 2^r for all values of r.

Tables 8 and 9 give the observed distributions of $\bar{x}$ and $\bar{y}$ from the three populations as well as the observed and expected values of the constants.

TABLE 8. FREQUENCY DISTRIBUTION OF THE MEANS.

central values	Population I		central values	Population II		central values	Population III	
	$\bar{x}$	$\bar{y}$		$\bar{x}$	$\bar{y}$		$\bar{x}$	$\bar{y}$
0 ·5	4	5	1 ·25	2	2	4 ·5	1	2
0 ·7	12	7	1 ·75	6	6	5 ·5	7	3
0 ·9	7	11	2 ·25	14	11	6 ·5	15	21
1 ·1	14	14	2 ·75	15	20	7 ·5	36	40
1 ·3	12	13	3 ·25	36	31	8 ·5	42	39
1 ·5	29	26	3 ·75	37	35	9 ·5	42	32
1 ·7	15	15	4 ·25	25	34	10 ·5	26	29
1 ·9	23	15	4 ·75	28	26	11 ·5	17	19
2 ·1	8	16	5 ·25	21	14	12 ·5	6	5
2 ·3	25	20	5 ·75	6	11	13 ·5	4	8
2 ·5	8	17	6 ·25	2	1	14 ·5	3	1
2 ·7	14	7	6 ·75	5	4	15 ·5	—	1
2 ·9	6	14	7 ·25	1	3	16 ·5	—	—
3 ·1	7	9	7 ·75	2	1	17 ·5	1	—
3 ·3	6	1	8 ·25	—	1			
4 ·5	4	3	8 ·75	—	—			
3 ·7	1	2						
3 ·9	3	1						
4 ·1	—	—						
4 ·3	1	1						
4 ·5	—	1						
above 4 ·5	1	2						
Total	200	200	Total	200	200	Total	200	200

Agreement of the first two moments with theory is everywhere perfectly good. The higher moments show deviations only in the case of the distributions of $\bar{x}$, but no explanation for this behaviour is evident from an inspection of the frequencies. Observed β_1 and β_2 give fairly good approximations to theoretical values in all the three cases.

TABLE 9. CONSTANTS IN THE DISTRIBUTION OF THE MEAN.

constants	Population I			Population II			Population III		
	observed		expected	observed		expected	observed		expected
	x	y		x	y		x	y	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mean	1 ·9500	1 ·7420	2 ·0000	3 ·5275	3 ·4975	4 ·0000	9 ·1250	9 ·1050	9 ·0000
μ_2	0 ·6686	0 ·6392	0 ·8000	1 ·3559	1 ·4867	1 ·6000	3 ·8211	3 ·8857	3 ·6000
μ_3	0 ·3506	0 ·5277	0 ·6400	0 ·6706	1 ·0051	1 ·2800	5 ·2403	3 ·7697	2 ·8800
μ_4	1 ·4401	2 ·1191	2 ·6880	4 ·5058	8 ·2453	9 ·2160	62 ·7680	45 ·7864	42 ·3360
β_1	0 ·4112	1 ·0062	0 ·8000	0 ·1804	0 ·3074	0 ·4000	0 ·4922	0 ·2422	0 ·1778
β_2	3 ·2217	5 ·1862	4 ·2000	2 ·4508	3 ·7304	3 ·6000	4 ·2989	3 ·0325	3 ·2667

SUMMARY AND CONCLUSIONS

In this paper an attempt is made to study by means of experimental sampling the distribution of r when the parent population is of the type III form and where the variates are highly correlated. For this purpose, 200 samples of five were taken from each of three different populations whose distribution laws are known, and having correlations 0 ·5, 0 ·75 and 0 ·89. The distributions of transformed r as well as those of variance and mean have also been recorded to see how far their first few moments and betas agree with theoretical results.

It has to be noted that skewness is maximum for Population I and minimum for Population III. To study the significant changes in the r-distribution due to changes in population skewness or population correlation, we should have taken either all populations with the same correlation and varying skewness or vice versa. Further it might be of interest to draw samples from populations with different marginal distributions. Such detailed analyses are, however, beyond the scope of this paper. It might be mentioned however that in a previous investigation of this nature (1941), the change in population skewness is found to have no apparent effect on the symmetry of the r-distribution when the population correlation is kept constant at zero. But it is observed here that just as in the normal case, the r-distruibution becomes more and more asymmetrical as ρ increases from 0 to 1.

Some of the main results emerging from this study can be stated as follows :—

(1) Close agreement with 'normal theory' frequencies is found only in the ranges 0·2 to 0·7, 0·1 to 0·8 and 0·3 to 0·8 respectively for the three distributions of r. At either end of the range, the total frequencies agree well with theory. For Population III, the frequencies at the lower end of the range do not show agreement either with 'normal theory' or the approximations.

(2) The first two sampling distributions of z show good agreement with 'normal theory' as is evident from the values of the frequency constants. In the third case, the only divergence from theory is for frequencies above 2·60. The observed distribution is slightly more leptokurtic than would be expected from 'normal theory'. In all three cases, the total frequencies on either side of the theoretical means are fairly equally balanced.

(3) Among the variance-distributions, good agreement with theory is noticed only for sampling distributions from Population II.

(4) All the mean-distributions show good agreement with theoretical values.

REFERENCES.

1. CHERIYAN K. C. (1941) A bivariate correlated gamma-type distribution function. *Journal of the Indian Mathematical Society.* (New Series) **5** (3). 133-144.

2. CHURCH, A. E. R. (1925) On the moments of the distribution of squared standard deviations for small samples of N drawn from an indefinitely large population. *Biometrika* **17**, 79.

3. FISHER, R. A. (1915) The frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10**, 507-521.

4. FISHER, R. A. (1921) On the probable error of a coefficient of correlation deduced from small samples. *Metron* **1** (4) 3-32.

5. HEY, G. B. (1928) A new method in experimental sampling illustrated on certain non-normal populations. *Biometrika* **30**, 68-80.

6. IRWIN, J. O. (1927) On the frequency distribution of means of samples from a population having any law of frequency with finite moments. *Biometrika* **19**, 228.

7. KRISHNAN NAIR, A. N. (1941) Distribution of "Student's" 't' and the correlation coefficient in samples from non-normal populations *Sankhyā* **5** (4), 383-400.

8. LE ROUX, J. M. (1931) A study of the distribution of the variance in small samples. *Biometrika* **23**, 134-190.

9. LEONE CHESIRE, ELENA OLDIS AND PEARSON, E. S. (1932) Further experiments on the sampling distribution of the correlation coefficient. *Journal of the American Statistical Association* **27**, 121-128.

10. PEARSON, E. S. (1929) Some notes on sampling tests with two variables. *Biometrika* **21**, 337-360.

11. PEARSON, E. S. (1932) The test of significance for the correlation coefficient : Some further results. *Journal of the American Statistical Association* **27**, 424-426.

12. RIDER, P. R (1932) On the distribution of the correlation coefficient in small samples. *Biometrika* **24**, 382-403.

13. SOPER, H. E. & OTHERS, (1915) On the distribution of the correlation coefficient in small samples. Appendix II. to the papers of 'Student' and R. A. Fisher—A co-operative study. *Biometrika* **11**, 328-413. Tables 379-413.

14. STUDENT (1908) On the probable error of a correlation coefficient. *Biometrika,* **6** 302.

15. TCHOUPROFF (1916) On the mathematical expectation of the moments of frequency distributions. *Biometrika* **12**, 193-194.

Paper received : 16 September, 1942

STATISTICS AND SOCIAL VALUES*

BY N. C. METHA I.C.S.

Lucknow.

While I am conscious of the great honour done to me by the invitation to address a meeting of specialists and learned men well versed in the technique of statistics, I am profoundly conscious of my deficiencies for the task allotted to me. I have, however, the little advantage of having been trained up as a student of economics at a time when the science of statistics was a part—although a minor part only—of economic studies. It was considered enough—I am speaking of the period just preceding World War I—to be in a position to recognise the obvious flaws in the statistical presentation of figures and also to learn the use of various indices. The last quarter of a century has ,however, wrought a veritable revolution in this field. A glance at the contents of the last number of the excellent periodical *Sankhya* would suffice to show the range and strides of modern statistics. Apart from the usual excursions in the higher reaches of mathematics, there are studies dealing with census, anthropometry, periodicity in human reproduction, examination of health data of university students, standardization of marks and even a factual analysis of aesthetic judgment.

Before I proceed to discuss the subject of this evening's discourse, I would like to pay a tribute to my old friend, P. C. Mahalanobis, who has been responsible for putting Indian talent on the statistical map of the world. The Statistical Institute has now attained international repute, and despite the fact that it has its branches in Mysore, Madras, Bombay, Lahore, Poona and now at Lucknow, the headquarters are still housed in the modest premises of Mahalanobis's own laboratory. His modesty is only matched by the magnitude of his achievements which are the fruit of almost single-handed labour during the last 10 years and more. The periodical that he has been sponsoring for the past 8 years would be a credit to scientific literature in any part of the world and I would like to take this opportunity of congratulating the galaxy of young men who have been responsible for maintaining such a high standard of scientific research in the pages of the *Sankhya*. I would also like to thank Prof. Radha Kamal Mukerji for readily accepting the task of organising a branch of the Statistical Institute at Lucknow and convening the first meeitng at such short notice. I hope that the branch institute will flourish under his enthusiastic guidance, for in these provinces we are particularly well-situated for an intensive study of statistics of all kinds. With five universities in our midst, there should be no lack of talent. It is, however, a study which requires not only hard work but also extreme precision and clarity of thought. There is hardly any room for vagueness of conception or slipshod handling of material.

With the expansion of industries the volume of statistical data is constantly increasing. This, however, does not necessarily mean that the numerical presentation of certain facts primarily prepared for administrative purposes makes us any wiser than before, unless special care is taken in defining exactly what it is proposed to achieve by the collection of such statistical data. Students of statistics in fact know that while there are mountains of figures to be found in almost all administrative offices, these mountains have little significance and are capable of yielding but little, primarily because the lay-out in the matter of collecting these data has not been scientifically designed. It is curious that while the necessity for statistics has been implicitly recognised by all Governments, the desirability of having a competent statistical staff to advise on the proper methods of collecting the data and utilizing them has been but very partially recognised in most of the states. The result has been that vast masses of statistical data are only significant of wasted labour. Fortunately, however, the outstanding results achieved by modern statistical technique have led to a growing recognition of the usefulness of proper statistical advice in the matter of laying out the experiments and interpreting them. The Imperial Council of Agricultural Research was perhaps the pioneer in this country in adopting modern methods of statistics and utilizing them in co-ordinating agricultural experiments laid out in various parts of the country. The result was, however, not achieved without considerable cost and also perhaps some expensive mistakes. It is, however, good to see that scientists in various branches of knowledge have now come to recognise that the part of their work, which is susceptible to numerical representation, is likely to gain in significance if it is handled by a qualified statistician, so far as the collection and interpretation of numerical data is concerned. It seems to me that it would be excellent if there were a Central Statistical Bureau at the headquarters of every Government whose duty it would be to see that the various figures collected for the various departments of the state were gathered in accordance with the accepted canons of statistics and were to be periodically studied and analysed. What is perhaps required is not so much the collection of more figures as the simplification of the methods used and the elimination of the obvious sources of error.

---

* Presidential address delivered at the inaugural session in Lucknow of the United Provinces Branch of the Indian Statistical Institute on 15th November 1941. This and the following papers by Radhakamal Mukherjee, Shitla Prasad Saxena, N. G. D. Joardar and J. K. Pande (which were presented before the same session) reached Calcutta between December 1941 and April 1942 in the midst of active evacuation of the Institute from Calcutta to Giridih, and had got mislaid. We are publishing these with apologies for the great delay.—*Editor*

I propose to refer to some of the difficulties which are inherent in interpreting the statistics dealing with social sciences, for one is apt to gloss over or to ignore altogether the background of concepts, conditions and ideals, which must be clearly understood before one can properly understand the new medical data. Let me, for instance, refer to the figures dealing with sickness and mortality. As one analyses elaborate tables of vital statistics, one is apt to overlook the tremendous social significance of births and deaths of which these figures are but outward symbols. No set of figures can possibly convey the significance in terms of human values; and it is precisely these values which must always be kept in the forefront whenever one is dealing with masses of figures. Take another instance of casualities arising as a result of the war. Here it is not a question of so many deaths but a vital matter determining not only the present but also the future composition of the entire population. And such change may not only affect the physical characteristics of the people but will also have immense psychological effects which may continue for a long time. For instance, it is said that as a result of the prolonged wars of the French Revolution the stature of an average Frenchman declined. There can be no doubt whatever that far-reaching effects in both the physical and the mental spheres were produced as a result of the holocaust of the great war, and the psychological make-up of the generation which came forward at the end of the great war was vitally affected by the tremendous mortalities suffered by the various countries in the world.

What I want to stress is that while the numerical results of several sets of figures may be the same, their significance may be quite different because the social or the conceptual background of these figures is different. Can there be any doubt, for instance, that a large casuality list directly as a result of hostilities is bound to have a very different set of reactions from an identical number of mortality arising out of natural causes ? One may lead to a psychological revolution, while the other may not produce even a ripple on the traditional surface of society.

Let me take another set of figures which in times like these come into prominence. We are all interested for instance in price indices, for they are not merely important from the administrative point of view but interest the manufacturer and the workman alike. The question of wages is, of course, organically connected with a correct working out, proper weighing and objective interpretation of price indices. A great deal of refinement has been introduced in the matter of working out these index numbers and considerable amount of care and objectivity is devoted to the proper distribution of weights so that the indices so prepared may be truly representative of the worker's actual budget of expenditure. But when all the refinements and subtleties of modern statistics have been exhausted, there still remains a substantial and a very important hinterland where only a social vision-imagination combined with sympathy will avail to light up the dark alleys of social life. Mathematical ingenuity will not lay bare the significance in terms of human values, for statistics are after all merely a means of presenting certain data illustrative of specific phases of life. Numerical or graphic presentation doubtless imparts a certain definition and precision to the ideas, but it must not be forgotten that they only show an outline of the picture and not the whole of it. Take, for instance, an ordinary price index number which we use in ascertaining the quantum of relief which should be given to the workmen. While weights are carefully distributed for various articles of necessity, no account is or can possibly be taken in the matter of certain vital necessities of life. For instance, a modicum of medical relief is required by everybody and a certain amount of money has to be provided for the purchase of medicines and drugs. Supposing the price of these articles goes up beyond a certain level as to drive the article beyond the purchasing capacity of the workman, or if the articles cease to be easily available at any price whatever, the impact of such a state of affairs on social welfare is out of all proportion to its numerical value in an itemised schedule of a workman's budget. Take even a simpler case. There is at present a definite shortatge of cloth and cloth prices have sharply moved up during the last few weeks. While a certain amount of economy in consumption is doubtless possible, there is a limit beyond which economy cannot go without serious repercussions on national welfare. A certain minimum in the matter of food as well as of clothing must be available if the people have to maintain their minimum standards of efficiency and living. What that particular figure is, it is at present not possible to say, but in a capitalistic society, particularly when Governments are involved in war, these considerations have to be kept in mind lest the national standards of living may be adversely affected as a result of abnormal strains and shortage of commodities.

Let me cite another example. We have recently increased the working hours of labour in order to increase the supplies of certain commodities. While the effect of longer working on the amount of the articles produced is obvious, it is not quite obvious what the effect is likely to be on the worker himself as a result of increased wages, enhanced strain, diminished leisure and decreased contact with his family. It is of course well-known that there is an optimum limit of working beyond which efficiency reacts inversely to the increased hours of working. But the question of efficiency in terms of output is not quite so important as the effect on the individual as a social unit. It is of particular importance if national efficiency is to be maintained at least at the existing level, if not at a higher level, that the increased strain of life must be compensated in such a fashion that the reduced leisure is intensively utilised so that the individual continues to function not only with his previous energy but also as a citizen who is able to discharge his duties with equal effectiveness *vis-a-vis* his family, his particular social group and the state.

In this connexion it would be interesting to note the practice followed in the army. Here the State immediately takes over the maintenance of the individual at the highest possible level of efficiency and therefore organises the food, clothing, medical relief and the use of leisure of the individual soldier. It goes even further and tries to relieve the men of any anxiety on the score of their family members by systematic and periodical work amongst them. In other words the army realises that the maintenance of efficiency is a function of a multiplicity of different factors and all these have to be carefully co-ordinated if the maximum result is to be achieved. The effect that soldiering produces under such circumstances is not merely confined to the period of the war but is far-reaching. It is by no means an accident that the Punjabi worker, skilled and unskilled, is found throughout the length and breadth of India, commands better wages and perhaps has a greater initiative and stamina than people of his own class outside his own province. The reason is that soldiering is at present the only occupation where the individual needs are considered as a whole and consequently the individual—I am primarily talking of men and not of officers—is deliberately organised and trained to the highest possible pitch of efficiency. An ex-soldier has a certain amount of initiative and *elan* which would not be ordinarily found among the poorer classes.

To carry the argument a little further. It is obvious that modern wars are as much conflicts waged by men as by machines and that the factory worker is as much needed on the industrial front as the soldier on the line of battle. If this is more than a mere piece of rhetoric, it is obvious that the organisation of the industrial worker falls very far short of the standard to which we have been accustomed in the fighting forces of the country. Apart from any question of social welfare, even from the point of view of immediate interest and requirements of the country, it is obvious that the organisation of the life of the industrial worker requires to be very much more intensive than is the case at present. If longer hours, better wages, larger total income, diminished leisure, increased strain, if all these are not to react adversely on the industrial mechanism, it is imperative that the worker should be looked after more or less on the same lines as the worker in the field of battle. For instance, it is important in the highest degree that increased wages must not be dissipated in drink or that increased strain must not result in avoidable breakdown or sickness. All this means the development of a new outlook as regards the organisation of our economic life, at any rate during the time of war. A certain amount of work is already being done by welfare centres but very much more is needed if the increased strain of working longer hours is to be borne without any deleterious effect on the workers themselves. The reforms obviously indicated are the establishment of community kitchens, greater facilities for medical relief and revolutionary changes in the matter of housing the bulk of our urban labour. These changes are needed to keep up and increase industrial efficiency. The point that I wish to make is that these are not questions of labour welfare so much as inevitable reforms if the industrial economy is to be maintained at its highest potential.

In this connexion I might cite rather an interesting episode from the history of mediaeval India. When the famous temples of Abu were being built(as acts of piety), it was discovered that the progress of work was slow. The question was investigated and it was found by the versatile lady Annapurna Devi, wife of the famous minister Tejpal, that a great deal of time was wasted by the workmen in keeping themselves warm in the morning on the hill tops of Abu and in preparing their meals in the afternoon, with the result that the working day was a very short one. The far-seeing woman immediately grasped the situation and organised lodging and boarding on a community basis and the progress of the work was thereafter found to be eminently satisfactory. This mediaeval example is very much to the point in dealing with the problems facing a great number of our urban workers. Their conditions of life cannot by any means be regarded as calculated to elicit their maximum endeavour. It is, therefore, all the more essential that careful planning and comprehensive examination of the various factors should be undertaken in the matter of planning new industry and organising the existing resources.

Before I leave this particular subject I might also say a few words regarding our cottage industries. The sugar industry in this country and particularly in these provinces has made enormous strides during the last few years. Luckily, however, for these provinces the sugar industry is still subordinate to the making of *gur* just as the making of *ghee* is still by far the more important method of processing milk than the supply of dairy milk in urban areas. The reason for this preference is obvious. From the point of view of the villager's dietary it is important that he should have his *gur* and *ghee*—or if *ghee* is too costly then his *matha* and skimmed milk. The *gur* that is available in the villages is vital for the maintenance of efficiency of men and animals in rural areas and it is just as well that we are still in a position to retain these products—primitive if you like to call them—still in the village, for the villager is not at present in a position to afford the costlier but the less nutritive articles used by the town-dweller.

May I also in passing refer to the various elaborate studies of rural life which have recently been published in all parts of the country? They all tell the same depressing tale of poverty, debt and the large expenditure of the income of the grower on food and clothing. There is, of course, some variation in the distribution of the resources among the various items of the budget but is it not desirable to break the routine of these enquiries and study samples of improving villages and ascertain the causes thereof? The economic map of the country is being constantly modified by various factors and the impact of these factors is one which requires careful analysis and study. Statistical investigators cannot afford to waste their time in continuing to produce almost standardised reports of this study of poverty in rural India.

When we come to consider questions of this nature the task of the statistician or the student of social science becomes enormously difficult; but the help of statistics is invaluable in dealing precisely with questions of this complexity. Statisticians in other words have to be more than mere scientists arranging and analysing masses of figures according to the latest methods of their science. They have to be social investigators and have to equip themselves with a vision which would penetrate and see beyond the arrays of their own symbols and read their trends in terms of human values.

Looking through the pages of the *Sankhya*, I am struck by the great diversity of the contributions. Recently, for instance, in an investigation among university students it was found that the incidence of skin diseases among girl students was 10 per cent as against 4 per cent among boys, though the condition of general well-being was more in favour of the girls than of boys except in the matter of wearing glasses. It was suggested that the higher incidence of skin diseases was due to the vanity of the gentler sex and lack of discrimination in the choice of soaps and cosmetics, similarly the larger use of spectacles was not, it was found, due to any particular feminine myopia so much as to the supposedly aesthetic value of the lenses.

Recently astrology has come into great vogue—particularly in Europe and America. We in this country have been trying to forget it as a mere superstition, but as a result of the unusual strain of the war this particular branch of human knowledge has come into considerable prominence and apparently has a certain amount of effect on the lives of the average citizens. Whatever the validity of astrology may be it is curious that statistics should have left it severely alone. It would be interesting, for example, to apply the modern statistical technique to some of the problems dealt by astrology. The most rational thing that I was ever told about astrology was by the late Rai Bahadur Vaidya—the historian of mediaeval India. He told me that it was extremely doubtful whether planets and stars had anything to do with the happenings of an individual life, but that it was not beyond the bounds of possibility to arrive at a reasonably correct prediction by merely examining a sufficienctly large sample of horoscopes. I should imagine that the method would be similar to that adopted by the insurance companies in arriving at the standard mortality figures. In any case the problem is worth investigation in view of the universal interest in this particular subject as well as the general gullibility of the average citizen. I do not know whether an astrologer like the finger expert makes the claim that no two individuals are ever born at precisely the same moment, for without some such assumption it would be quite impossible to account for a solitary Buddha, Christ or Mohammad or Napoleon or even Hitler at a particular period of history.

Leaving astrology aside, let me turn to the repercussions of the recent rise in income-tax. While the State has obviously been getting a larger and larger share of the war time profits it is perhaps not realised that the owner or the manufacturer or the managing agent has also been getting a larger proportion of the share of the industrial income than before, for his share of the income is generally calculated before anything is allowed for the payment of income-tax or for the depreciation of the plant. With a higher incidence of income-tax it is obvious that the proportion going to the share-holder has diminished, while that of the managing agent has increased. Whether this is equitable or not is a matter with which I am concerned. I only want to point out the shift in the distribution of income which is taking place as a result of the change in the scale of income-tax and the imposition of the excess profits tax. It is perhaps needless to say that these variations in the distribution of income are only incidental and were perhaps not contemplated at all by the exchequer.

I have taken these examples merely to illustrate the vastness of the field that lies for the student of statistics. The modern technique is economical and does not wholly rely upon masses of departmental statisics but with the complexities of modern life it is more necessary than ever before to keep in prominence the aim of all our social activities—namely the promotion of social welfare. Unless this is done, the imposing facade of statistical analysis is apt to be merely a piece of intellectual vanity.

Paper received : 12 December, 1941.

STATISTICS IN THE SERVICE OF PLANNING.

By RADHAKAMAL MUKHERJEE
University of Lucknow

The movement of the social sciences to-day is towards greater objectivity and quantitative precision. The aim is not only to give a quantitative expression to human affairs but also to plan and predict them. Statistics with its special techniques and methods for handling social data quantitatively thus becomes an indispensable logical apparatus in these tasks of the social sciences.

More significant than the theoretical import of statistical studies is their role in social and economic planning which is demanding attention from social scientists throughout the world. It will be worth while to indicate here the major statistical criteria which should underlie the social and economic planning we are envisaging for India and by which we can measure planned advance from year to year. With the greater part of the world at war, it is apparent that man's control of physical forces has far outstripped his capacity and/or will to plan and direct human affairs. To plan society certain objective tests of progress have to be assumed and worked upon. The general test, of course, is the improvement of the standard of living of the masses. But it is necessary to envisage certain objective living optima or standards which may be reached within a period of, say, 5 or 10 years in the country. The National Planning Committee[1] at its second session appointed an Aims and Purposes Sub-committee, (of which I was Chairman), which laid down certain statistical tests for measuring the progress of an economic plan in India. These were :

(1) The improvement of nutrition on the basis of irreducible minimum requirements of proteins, carbohydrates and minerals (as well as necessary protective foods) from about 2400 to 2800 calories for an adult worker.
(2) The improvement of clothing from the present consumption of about 15 yards on an average to at least 30 yards per capita per annum.
(3) Housing standards to reach at least 100 sq. ft. per capita.
(4) The improvement of literacy from about 10 per cent to 100 per cent.
(5) The improvement of the average expectation of life from 26 to 50 years.

The Anglo-American school of economic thought formulates a narrow atomistic theory of standard of living. Continental writers in the west have developed a much wider conception of living standards and speak of "the totality of an individual or group's state of felicity which should be co-ordinated with the entire living condition". Environment in its broadest sense is to be co-ordinated with standards which include working hours, rate of accident, morbidity and mortality rates, as well as housing, food and clothing, leisure and recreation. Such an ensemble cannot be expressed quantitatively by one single index of progress. On the whole, however, the increase of the average income and increase in the average expectation of life may be taken as excellent tests of economic progress and, indeed, the combination of these two tests may measure both economic and many non-economic cultural elements in social life and welfare. To establish a national minimum income and standard of health and economic security is accordingly the primary task of planning. War time, it is said, is not opportune for social and economic reform in India. It must be pointed out, however, that in England and in America the basic economic reforms sprang out of war conditions. The Atlantic Charter laid down the maintenance of national standards of living and social security as fundamental peace aims and the right of all countries. In India we must now establish minimum wages for unskilled workers and for lower grades of technical workers and introduce a scheme of insurance, to cover the risks of sickness and invalidity other than those covered by the Workmen's Compensation Act. Many years back Britain established Trade Boards to set wages in low paid industries while the social services that have now been organised have for their aim the maintenance of the worker's standard of living when he is a victim of unemployment, sickness or old age. The major war industries in India are iron and engineering. In most countries of the world the average wage in iron and steel industry is double that of an ordinary industrial worker. In India the wage in iron and steel and engineering industry is actually lower than the rate prevailing in the textile industry.

---

[1] Appointed by the Indian National Congress in 1939.

In the biggest engineering establishment in India about 12,000 workers or 40 per cent of the employees do not obtain the living wage[2] of Rs. 20/- per month worked out on the basis of an irreducible scale and cost of living. In all Government contracts during the war there should be a compulsory clause safeguarding to workers the payment of minimum wages and sickness leave and allowances. War time planning dovetails into peace-time planning. On the other hand, the accumulation of profits of companies and industrialists without safeguarding minimum wages and standards of living is bound to lead to a series of strikes like those which followed the last European war. It is also sure to create an embittered atmosphere that is incompatible even with economic stability what to speak of economic planning.

The demand for the industrial workers' minimum wages and economic security affects, however, only 5 millions of Indian workers employed in the factories out of a total of about 154 million workers in the whole country. They are mostly agriculturists, artisans, traders and professional classes. A marked improvement of the standard of living of industrial workers and urban dwellers in India cannot be reached without the improvement of conditions in the villages. In order to counteract the vicious circle of subsistence farming, low standard of living, poor purchasing power, and industrial backwardness, the attack should simultaneously begin from the agricultural front with improvement of farming methods, practices and organisation and from the industrial front with the stress on rural industrialisation.

Statistics in the service of economic planning reveals the necessity of integration and comparison of activities in different sectors of social and economic life. The improvement of agricultural production and the speeding up of industrialisation in India—which we may increase the average income and the span of life of the people—mutually support each other. An increase of income of cultivators can effectively speed up industrialisation and at the same time slacken the present rate of multiplication. Industrialisation specially in the country-side which may use the primary raw materials from agriculture can mitigate agricultural unemployment, and improve both agricultural efficiency and income. At the same time the engendering of new wants among the cultivators that rural industrialisation fosters would have far reaching social effects. With industrialisation, again, the surplus labour in the fields and farms would be deflected to workshops and factories and make tenancy reform practicable. Planned advance is accordingly cumulative—a co-ordinated advance on all fronts. For measuring such co-ordinated planned advance, statistics should, therefore, work on the basis of certain synthetic indices of progress : for instance, the correlation of expansion of cottage and agricultural industries with the decrease of small tenants and the floating landless class ; the correlation of increase in agricultural yields and subsidiary employment with increase in purchasing power and decrease in birth-rate ; the correlation of increase in wages and standard of living with increase in productivity of labour and decrease in cost of production, and so on. The Gosplan in Soviet Russia uses such general "balances" for shaping and appraising the plans. Statistical schedules prepared by each economic unit of the country will represent the exact inductive basis of the accounting, checking and control of economic planning.

---

[2] At the level of 1941 prices.

But the Nation's social optimum cannot be wholly expressed only by high average income, longevity, employment and productivity, though these latter have the merit of measurability that is denied to the social or ethical optimum. Sociology is, however, now developing tests for balanced progress. Those which have proved useful for the analysis of social trends and direction of social policies are : minimum differences in attitudes, opinions, or scales of preference between groups or classes with appreciable differences in income, maximum unison of public opinion or unanimity of groups or political parties, maximum recognition of individual rights as judged by rarity of imprisonments and other infringements of liberties, etc.

The social optimum is integral. A community with a high average income but with less leisure and freedom in comparison with its standards at an earlier date or with those in other regions belonging to the same order of civilisation may be considered to have diverged from optimum conditions. In India the increasing disagreement between political parties and the social cleavage, between the different classes, castes and communities are as much signs of social backwardness as poverty, unemployment, inefficiency of man-power and low average longevity. It is necessary, therefore to examine changes in individual standards of living and standards of public welfare, uses of leisure as well as changes in the economic rights of workers and general social and political stability. These can be given quantitative expression only partially, and yet without them we have to abandon the integral optimum concept. Statistical studies of the use of the vote and of attitudes and opinions of different classes in the society segregated by differences in income and material-cultural equipment, for example, should reveal the extent of social and economic cleavage or solidarity. Every province and every district in India working under a National Planning Commission should therefore have arrangements for making planning schedules by which each unit will be able to measure the stability of its tenancy and improvement of agricultural yields, the increase of horse-power in industrialisation and decrease in cost of production as well as the improvement of average income and longevity, the improvement of public welfare and the abolition of social, economic and political cleavages.

Paper received : 13 April, 1942

DIFFERENTIAL RATES OF NATURAL INCREASE.

By N. G. D. JOARDAR.

Lucknow

The growth of a community is attributed to three factors : (a) natural increase or the balance of births over deaths, (b) excess of immigration over emigration, and (c) fresh addition through change of religious affiliation. While the last two sources of growth are constant, the first operates in all groups under all normal circumstances. In an expanding community more births take place than death ; in a steady community, the birth rate is offset by the death rate, but in a declining community, deaths make more gaps in the line than births can fill.

Present enquiry. Six communities in the city of Lucknow were studied : the Indian Christians, Bengali Hindus, Pahari Hindus, Khattris, Kashmiri Pundits and Sunni Muslims. The first census was taken on January 7, 1938 and was repeated in the three successive years. The numbers for the different communities living in family organization, their births in the calender years 1938, 1939 and 1940, the birth-rate expressed in percentage and the natural increase per year as obtained from the census are tabulated in cols. (2), (5), (8) and (11) of Table 1.

TABLE 1. NATURAL INCREASE IN DIFFERENT COMMUNITIES.

name of community (1)	number (2)	males (3)	females (4)	births in different communities (5)	no. of males within reproductive period (6)	no. of females within reproductive period (7)	available women within reproductive period per 100 males (8)	birth-rate per cent per year (9)	mortality rate per cent per year (10)	natural increase per year (11)
Sunnis	5554	2978	2576	419	1423	1254	88 ·1	2 ·51	1 ·40	1 ·11
Khattris	4556	2186	2370	354	1123	1234	105 ·8	2 ·58	1 ·33	1 ·25
Paharis	4398	2232	2166	295	1204	1157	91 ·3	2 ·22	1 ·11	1 ·11
Indian Christians	4032	2000	2032	284	905	941	104 ·4	2 ·36	1 ·22	1 ·14
Bengalis	4025	2001	2024	296	900	942	100 ·0	2 ·11	1 ·11	1 ·00
Kashmiris	660	335	325	39	160	165	100 ·0	1 ·97	0 ·62	1 ·25

Factors in growth. Of the six groups of people studied, the birth-rate is the lowest among the Kashmiris ; yet they have the highest rate of natural increase. The other communities with larger birth- rates have smaller survival rates. There were four definite factors to account for natural increase in these communities. The positive factors were (1) the availability of adult women in the reproductive age, (2) marriage in the most productive period of life, (3) a solvent economic life and happy adjustment of environment, and (4) a sudden release from the old modes of a community life. These four factors were modified in various degrees by four other ones which acted as checks to the natural growth of the communities. These were (1) poverty, diseases and ignorance, (2) mortality rates, (3) a changing conception of family life and (4) society and public life.

Availability of adult women in the reproductive stage. It is a common place that adult men and women are the potential parents of children and authors of natural increase. But the period of productivity is not uniform in all countries. From the present survey it was found that the reproductive period lies between 16 and 43 years in case of women and between 19 and 46 years in case of men with a negligible margin of duration here and there. Taking these as the reproductive periods, the number of males and females within the periods were found and are shown in cols. (6) and (7) of Table 1. But all these women were not available for reproduction. In the Khattri, Pahari, Bengali and Kashmiri groups, there were 95, 57, 87 and 9 widows respectively who for religious and social reasons could not marry again. Considering that 20 years was the average period for child-bearing counted from the time of marriage these widows had an average of ten years of married life. So their availability was reduced by half. The number of available women per 100 men after this reduction is shown in col. (8) of Table 1.

Marriage in the most productive period of life. One of the causes of high *fertility* is supposed to be marriage in the earlier part of life because an early marriage brings with it a longer span for child-bearing and a late marriage a shorter one. From 2,156 studies of women of these six different communities, the average periods of production were found for women starting to bear children at different ages and are shown in Table 2.

TABLE 2. AVERAGE PERIOD OF PRODUCTION FOR WOMEN STARTING TO REAR CHILDREN AT DIFFERENT AGES.

age in years	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
average period in years of production	15·6	15·9	17·9	18·9	20·6	19·5	19·0	18·0	15·0	14·8	12·0	11·6	11·6	9·3	8·0	5·0	3·7
Total ..	31·6	32·9	35·9	37·9	40·6	40·5	41·0	41·0	39·0	39·8	38·0	38·6	39·6	88·3	38·0	36·0	35·7

From the stand point of natural increase the period between 20 and 23 years seems to be the most productive. So marriage between 18 and 24 years is a definite source of natural increase. It appears from the data that the Khattri community with its larger number of women and its judicious system of marriage in the most productive period attains its highest birth rate.

A solvent economic life and happy adjustment of environment. The birth of children was not proportionate to the economic condition of the parents even on an average. Richer families which had means to support a larger numbers of children had scarcely more children than the middle class families. It is remarkable to find that the poorer classes of all these communities have on an average fewer children. It seems that there is a correspondence between an optimum income and a larger number of children. Taking all the rich people of these groups together and comparing them with all the milddle-class people I found 31 per cent of the rich parents have 4·3 children per married couple while 32 percent of the middle class parents have 3·8 children. The conclusion is inescapable that income, though a factor by itself, needed to be weighted by other considerations.

A sudden release from old modes. A large number of poor parents in the Indian Christian community had much larger families than other people of the same category. These parents one and all, were converts from the Hindu or Muslim faith into Christianity.

Was change of religion, then, a stimulus to natural increase? The figures at our disposal were only for 142 parents. We can not come to any conclusion. But the data collected in three large villages in the district of Nadia corroborated our hypothesis. Compared with their brothers and cousins, the converts had larger families, economic resources remaining the same. But in the second generation, the children of the converts definitely showed a pull in the other direction. They had fewer children than usual. From the study of 1,523 sons of converts, it was found that they had ·87 children less than other people of their class. The sons of converts had fewer children in most cases, not because they were sons of converts but for the fact that they were mal-adjusted to their environment.

Poverty, disease and ignorance. Poverty which freezes the sexual impulse and acts as a blight was fairly common among the lower strata of the Indian Chirstian, Sunni, and to some extent, among the lower races of the Pahari and Khattri communities. The Kashmiri and the Bengali communities were free from it. Eleven per cent of 675 Indian Christian and Sunni families suffered from it.

Diseases which are generally known as deficiency diseases, such as skin troubles, rickets, tuberculosis were fairly common among these poorer folks. Gonorrhoea, syphillis, neo-natal troubles are also noticeable among these classes. Poverty also led to pre-disposition to malaria and other diseases by sapping the vitality of the people and actually stood in the way of medical aid, proper hospitalization and suitable nursing.

Mortality rates. The mortality rates of the different communities are shown in col. (10) of Table 1. The Kashmiris due to wealth and education have reduced their rate of mortalty to the lowest possible figure which is comparable to the rates of New Zealand and Sweden.

A new conception of family life. A new conception of family life has come to the people in these communities who have received eductaion on western lines. Parents have come to think in terms of smaller families and this has acted adversely on the birth-rates of the communities. The educated middle and upper classes in all the communities, and almost the whole of the Kashmiri group, have caught the new spirit and we see signs of it in the falling birth-rates of these people.

Society and public life. Society and public life make demands on the time and energy of the people who get into them. Women especially find child-bearing to be an obstacle to their career. It is natural that contraceptive measures should come into vogue among these women as well as among those who want to avoid larger families. How widely the contraceptive measures are used in the different communities is not known. But there are reasons to believe that the practice is wide-spread where the western type of education has deeply penetrated.

Paper received: 13 April, 1942.

COST OF LIVING INDICES AT CAWNPORE

By SHITLA PRASAD SAKSENA
Lucknow

Introduction. This paper gives the results of an investigation into cost of living for Cawnpore industrial labour for the period 1914 to 1938. Cost of living index numbers have naturally a purely relative character and show the movement in the cost of living over a given period compared to a base period.

It is impossible to include all commodities in an index number but important commodities should be weighted according to their importance. Out of a number of methods of allotting weights the aggregate consumption and the family budget methods are mostly in use. The aggregate consumption method is more suitable for indices calculated for a region or a country as a whole but when the object is to compare the movement of cost of living indices and wages, or in other words, to measure the purchasing power of money for a particular class or community, the family budget method is the best.

A word may be said here about the collection of price statistics. In India, general index numbers of wholesale prices are published annually but this series is unsuitable for this inquiry. It may serve as a record of trade movement but it is of little value for measuring the purchasing power of money for any community, because the figures are expressed as percentages of the prices for the year 1873 ; the list of commodities has not been revised since 1889 ; it includes many commodities which are not important and excludes others and is unweighted.

For this investigation, information as to the predominant retail price, in various years, of cereals, pulses and some of the articles of food other than cereals and pulses as far as available was collected from the Government Gazettes published every week. Other prices were taken from the declared value of exports and imports of India as published annually in the Sea-Borne Trade of British India, and the Indian Trade Journal issued by the Director of Commercial Intelligence, Calcutta. Price variations in cloth were gathered from prices quoted in the Review of the Trade of India and the Sea-Borne Trade of British India. Reliable records of house rent in the city were not available ; where the labourer had been occupying the same house for a period of twenty years. In many cases, a rise in rent was preceded by an increase in accommodation or the comforts in the house. Figures for rent were collected from private *bustees* or *hatas* and the quarters in settlements provided by the factories.

Charges paid for various services do affect the budget and the living costs of the labourers but a study of the variations under these heads is not possible due to the absence of data and these items had to be ignored. In all, only about 73·3 per cent of the total expenses of the labourers could be covered by this inquiry.

Wages. Calculations of the movements of industrial wages are still difficult. It is not possible to obtain authentic published information on any large scale regarding the earnings of the industrial labour, the amount received by an industrial labourer per week or per month. This is due to the too frequent changes in the personnel of factory employees and to the absence of any record maintained for public information by the employers or the newly organised trade unions. In these circumstances an inquiry into the incomes can only be made through the labourers. Though the average labourer does not have the ability to understand the system of wage payment, he can at any rate mention the amount which he received as wages in a particular fortnight or month. Collection of data regarding wages through this source, no doubt, has its own limitations reports of various inquiries—official and non-official, serve as good correctives for the data collected first hand from these sources.

Cost of living. A change in the cost of living may be due to (1) a change in the purchasing power of money, (2) a change in the commodities consumed, and (3) to a combination of the above two factors. A cost of living index deals with the variations caused by the prices alone. In other words, it measures the change in the purchasing power of money. The particular commodities and the quantities of those commodities consumed were assumed to be the same in the subsequent years as in the base period. A change in the commodities consumed may be due to a change in tastes and fashion, or it may be due to a rising price level for the commodity which makes the use of the article prohibitive for the poorer classes.

At Cawnpore out of 729 family budgets collected by a process of random sampling there were 80·2 per cent Mohammadans and 1·6 per cent Christians. But for minor differences, due to customs and habits, budgets of the workers in the same economic grade from all the communities were more or less of the same type and showed the same peculiarities. In the same way, among the Hindus, differences in the diet of the workers belonging to different sub-castes of the Hindus were at times noticeable ; but in the present enquiry, to avoid minute classification and sub-classification, such differences were ignored.

Disparity in the budgets of the workers was studied purely on the basis of income grades, and an attempt was made to give due weight to the various economic grades among the industrial workers in the city.

Most of the Hindu workers were living on vegetarian diet, and hence a vegetarian diet was taken to be the basis for enquiry. Due to the absence of any published price statistics or articles like meat, fish, sweetmeats, milk, vegetables, fruits and the like, it was not possible to calculate the variations in the price indices of these commodities, and in this enquiry these articles were left out from the budget of the labourer.

Food was sub-divided into three distinct heads : (a) cereals, (b) pulses, and (c) articles of food other than cereals and pulses. As far as available, retail prices have been collected because the workers rarely make wholesale purchases. In the case of articles of food other than cereals and pulses such as butter, sugar, salt, linseed, oil, spices, etc. retail prices for these commodities are not quoted in any reliable publication. Moreover, in an enquiry where the movement of prices for various articles is reckoned in terms of indices, retail or wholesale prices are not likely to make much difference provided that for a commodity the same prices, retail or wholesale, are taken for the whole period.

Weights of food articles. On the basis explained above the index numbers of retail prices for different articles of food and their weights have been calculated for the years 1914 to 1938 in Table 1.

With a view to giving due importance to the variations in the incomes of the labourers, all workers have been classified into four grades of monthly income, namely, Rs. 15/- and below; between Rs. 15/-;and Rs. 30/-, between Rs. 30/- and Rs. 40/-; and between Rs. 40/- and Rs. 50/- ; separate weights are allotted to each grade of labourers on the basis of their actual monthly expenditure, and then over-all weights are calculated according to the number of labourers in each. The index numbers of the commodities for various years are now multiplied by the weights allotted to these commodities in each centre respectively and the products of the index numbers and weights are totalled and divided by 100, the sum of all weights shown in col. 12 in Table 1.

TABLE 1. PRICES AND WEIGHTS OF FOOD-STUFFS. (1909-1913=100)

weights	17·0	42·2	2·0	12·0	6·9	10·3	1·4	2·6	1·4	4·2	100·0
year	rice	wheat	gram	dal arhar	mustard oil	ghee	raw sugar	refined sugar	salt	spices	indices
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1914	113	122	163	166	107	120	96	93	100	105	123
1915	120	138	146	172	99	115	122	107	180	103	132
1916	128	116	125	147	98	114	113	147	240	118	123
1917	120	122	125	119	124	120	116	132	270	118	121
1918	137	154	163	147	127	135	96	209	330	128	149
1919	174	195	283	306	282	158	151	267	270	152	209
1920	207	176	221	291	167	197	192	217	310	131	199
1921	180	203	254	263	154	166	227	287	319	118	200
1922	167	197	221	197	166	168	178	275	200	133	187
1923	143	138	121	156	155	188	122	174	360	157	152
1924	150	138	92	156	156	186	127	161	300	181	152
1925	161	170	167	169	165	181	156	166	260	194	171
1926	170	170	184	213	168	171	162	181	250	182	177
1927	170	149	188	244	165	177	111	191	250	158	171
1928	172	151	188	234	151	172	118	172	250	158	169
1929	178	145	233	244	144	170	145	181	250	175	173
1930	140	100	163	191	145	171	124	244	260	155	138
1931	98	73	96	141	108	233	89	166	250	109	111
1932	102	92	96	141	92	144	78	127	250	108	109
1933	98	89	104	125	74	127	55	164	150	94	101
1934	87	78	96	122	65	97	78	166	270	96	92
1935	100	84	100	156	84	100	96	162	270	92	103
1936	107	87	94	150	100	98	68	114	290	92	104
1937	107	111	113	153	88	102	71	36	270	112	113
1938	102	84	108	159	82	101	85	29	270	111	101
1939	148	89	140	159	82	104	139	20	270	112	112
1940	106	109	146	141	—	—	89	—	290	—	—

Light and fuel. In calculating the indices for the cost of light and fuel, prices of kerosene oil, castor oil, coke and firewood have been taken into account. Table 2 shows the index numbers of prices and weights for these commodities.

Clothing. There are wide differences in the dresses of the labourers. The quality, size and variety of cloth change very frequently. It is therefore impossible to find out with accuracy the movement in the index number of prices of cloth. Nor are the records of the price fluctuations for all varieties of cloth kept and maintained. To get a rough idea of the general trend of prices a few well-known varieties of cloth which are standardised and for which regular quotations are available have been taken into account. Table 3 gives the index number for these varieties of cloth.

House rent. Workers reside in *bustees* and labour quarters provided by private landlords, public bodies and the employers. No regular rent records are available for the paid in the *bustees* and the quarters provided by private landlords. Workers themselves cannot exactly tell the changes in rent during the whole period of inquiry due to frequent changes of their tenements. At times variations in house rent mentioned by the workers are not very reliable. In many cases, an increase in accommodation or other comforts in a house brings about a rise in rent which, strictly speaking, cannot be treated as an enhancement. It is a compensation for additional comfort. Public bodies at Cawnpore have now devoted their attention to this problem and in some cases have even provided some labour quarters. But they serve no useful purpose for recording the movement in rent in this industrial centre. They are too recent and do not provide a basis for long-term comparison. They are also very limited in number and form a very small fraction of the total number of houses occupied by industrial labours. Only the mill settlements are thus left for comparison. House rent does not change from year to year. Compared to pre-war period, there was a rise in rent in 1920-21 when the war conditions caused a general rise in all prices. The rent index stood at 200 in that year (1913-14=100). Since then rent recorded practically no change till 1927-28. In 1927-28 there was again a rise, and these rates continue up to this day, the index standing at 381. There might have been minor changes now and then, but they do not vitiate the general trend.

Table 2. Prices and weights of fuel and light at Cawnpore. (1909—1913=100)

weights	13·6	2·3	—	78·1		Weights	13·6	2·3	—	78·1	
year	kerosene oil	castor oil	coke	fire-wood	annual weights	year	kerosene oil	castor oil	coke	fire-wood	annual weights
(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
1914	103	104	87	100	101	1927	158	159	155	147	150
1915	108	96	111	118	115	1928	135	165	138	151	150
1916	125	109	105	94	100	1929	138	156	133	142	143
1917	140	124	109	147	144	1930	138	147	134	125	129
1918	170	157	120	136	142	1931	133	124	137	113	117
1919	178	282	332	142	159	1932	115	111	138	104	106
1920	183	200	190	172	176	1933	95	103	130	94	95
1921	188	156	336	200	195	1934	93	107	119	79	83
1922	178	176	405	142	150	1935	93	113	91	83	87
1923	163	196	249	151	155	1936	78	111	120	87	88
1924	163	220	208	151	158	1937	78	111	130	94	93
1925	163	203	176	158	162	1938	93	119	110	97	98
1926	168	160	204	151	154	1939	93	134	132	97	100

Miscellaneous expenditure. There are a number of miscellaneous expenses incurred by the labourers every month. Items included under this head are : barber, dhobi, sweeper, medicines, education, railway fares, tobacco, liquor, toddy, pan supari (betel nuts), amusements, festivals, remittances abroad, interest on debt, repayment of debt, subscription etc. The most common and more or less permanent miscellaneous expenditure is on smoking and on the use of finely cut betel nuts for chewing. Tobacco leaf mixed with lime and betal nut chips is the commonest form of the use of tobacco, though indigenous smoking pipe known as *huqqah* and *biris* are also used. Cigarettes are less popular due to high costs. No regular records of the cost of other services and articles, which are casually used, are available and they have been ignored in the present enquiry. *Pansupari* and tobacco are equally important and hence equal weights are allotted to them. The following are the index numbers for betel nut and tobacco leaf and weighted index number for the cost of miscellaneous items.

TABLE 3. INDEX NUMBER OF CLOTH AT CAWNPORE (1909-13=100)

year	imports of cotton piece goods.			exports of cotton price goods			average
	grey	white	coloured	grey	white	coloured	
(1)	(2·1)	(2·2)	(2·3)	(3·1)	(3·2)	(3·3)	(4)
1914	107	106	105	107	84	158	111
1915	100	100	115	93	64	163	106
1916	127	118	150	100	55	142	115
1917	167	153	190	127	61	153	142
1918	240	224	225	187	89	211	196
1919	273	282	290	253	95	242	239
1920	307	312	355	300	139	279	282
1921	247	247	300	280	132	253	243
1922	220	224	255	247	109	253	218
1923	213	218	255	227	86	221	203
1924	227	224	255	207	95	216	204
1925	213	206	220	167	86	211	184
1926	180	188	200	207	89	205	178
1927	167	165	170	200	84	200	164
1928	160	165	175	193	84	195	162
1929	153	171	160	200	82	189	159
1930	133	147	145	167	95	184	145
1931	107	112	115	180	100	168	130
1932	100	106	100	180	66	168	120
1933	87	106	100	167	70	158	115
1934	73	112	85	160	59	153	107
1935	93	112	80	140	48	163	106
1936	87	118	95	320	109	263	178
1937	100	124	110	327	139	358	177
1938	80	112	100	327	157	389	209
1939	73	118	90	220	57	174	121

TABLE 4. MISCELLANEOUS EXPENDITURE AT CAWNPORE. (1909-13=100)

year	betel nut	tobacco leaf	weighted Index Nos.	year	betel nut	tobacco leaf	weighted Index Nos
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1914	180	143	162	1927	168	211	190
1915	181	110	146	1928	142	171	157
1916	90	107	99	1929	134	186	160
1917	89	132	111	1930	132	158	145
1918	97	155	126	1931	111	118	115
1919	111	131	121	1932	91	68	80
1920	123	146	135	1933	74	49	62
1921	113	162	138	1934	66	56	61
1922	119	188	154	1935	61	76	69
1923	118	161	140	1936	64	75	70
1924	129	149	239	1937	74	80	77
1925	154	180	167	1938	86	86	86
1926	171	202	186	1939	82	90	86

GENERAL COST OF LIVING INDEX

A weighted general cost of living index number is now calculated by multiplying these weighted indices for separate groups in each centre by weights for each group at that centre in the total expenditure in all groups and divided by the total number of weights for all groups. Table 5 shows the weighted general cost of living index numbers.

Table 5. Cost of living index at Cawnpore (1919-13=100)

year	index no.	year	index no.	year	index no.	year	index no.
(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
1914	131	1921	202	1928	165	1934	127
1915	137	1922	188	1929	204	1935	135
1916	129	1923	164	1930	169	1936	143
1917	135	1924	167	1931	147	1937	150
1918	159	1925	175	1932	143	1938	146
1919	204	1926	188	1933	135	1939	144
1920	204	1927	182				

The following are the weights for each group of expenditure : Food 48·2, Fuel-Light 6·2 ; Clothing 7·6, Rent 9·0 and Miscellaneous 2·4.

Paper received : 13 April, 1942.

THE CAWNPORE MILL-WORKER, HIS FAMILY AND HIS EXPENDITURE.

By J. K. PANDE, M.A.
Lucknow

Introduction. The present study is based on a preliminary analysis of 300 out of about 1400 family budgets of mill-workers collected at Cawnpore during the winter of 1938-39 under the guidance and supervision of the Bureau of Economic Intelligence, U. P. The statements made and the conclusions drawn in this study are, therefore, provisional and subject to revision and change in the light of facts which might be revealed by an analysis of all the budgets collected. Of the 300 workers whose budgets have been analysed, only 10 per cent were born in Cawnpore city while the rest came from villages, another 12 per cent belonging to villages in Cawnpore district itself. About 20 per cent came from Partabgarh, 13 per cent from Rae Bareli, 8 per cent from Unao, 5 per cent from Fatehpur and the rest from other districts in U. P. as far away as Gorakhpur in the east and Agra in the west. Only 6 per cent were drawn from districts situated west of Cawnpore, these being Agra, Etawah, Farrukhabad and Hardoi. A possible inference from this strikingly heavy leaning of the Cawnpore mill-worker towards the eastern districts of the province would seem to be that there is a greater pressure on land in the eastern parts of the Province.

Till a few years back it was commonly supposed, perhaps with some justification, that the Cawnpore mill-worker was principally an agriculturist and came down to Cawnpore in the off-agricultural season with a view to keeping himself more gainfully employed than would be possible in his village. The data under examination throw an interesting light on the subject, and reveal that the Cawnpore mill-worker began learning to settle at Cawnpore about 10 years back. Of the 300 families, 92 per cent have been in regular employment at Cawnpore for an average period of about 10 years since the time of their migration to Cawnpore. The number of breaks in employment per worker during these 10 years comes to an average of 1·5, of which about 1·0 were due to involuntary personal reasons, like the death of a senior member of the family at home necessitating a brief recall to the village, and 0·5 were due to reasons beyond the mill-worker's control, like unemployment.

The worker was also asked about the cause of his migration to Cawnpore; but the results given below have to be understood in the light of the proverbial incapacity of the U.P. villager to distinguish the immediate from the ultimate cause. In about 37 per cent of the cases the cause of migration was given as increased pressure on holdings resulting in surplus labour and in about 30 per cent cases economic ambition or adventure. In about 15 per cent cases the cause was assigned to poverty or lack of subsistence and in about 8 per cent to ejectment from holdings. In the remaining cases the causes were indebtedness, economic calamities, zamindars' oppressions or family quarrels. A note-worthy feature of these returns is that in a large majority of cases, to be exact in 96 per cent, the cause of migration given by the worker was economic.

The mill-worker's age. The average age of the Cawnpore mill-worker works out to 33 years, which read with his 10 years' stay at Cawnpore, gives his age at the time of migrating from his village as 23. This perhaps may mean that the average mill-worker does not like to start his career as a mill-worker in preference to his agricultural occupation, but forced either by poverty or by under-employment or perhaps urged by economic ambition so rare in a villager, he later has to come down to Cawnpore.

Family size and composition. The average number of persons per family of a mill-worker comes to 5·13 of which 1·78 are under 15 years and 0·77 over 50 years, 2·58 being between the ages of 15 and 50. The largest number of persons per family comes to 8·7 in families earning over Rs. 60 p.m. and the lowest to 4·1 in families earning less than Rs. 10. The number varies directly with the amount of income. An average family contains 2·76 males and 2·36 females.

Earners and dependants. Of the 5·1 persons in a family 1·2 are workers, 1·9 dependants living with the worker and 2·0 dependants of the worker living away from him at the village home.

Natural families and joint households. Of the 300 families under examination, 95 were natural families consisting of a marrried couple, their children and step-children, while 205 were joint households. The average number of persons in a natural family came to about 4 and in a joint household to about 5·6, the number of workers being about the same in both, which indicates, that *ceteris paribus*, the joint households were economically worse than the natural families. It is, however, interesting to note that the number of resident dependants was also about equal in the two types of families, so that the larger size of the joint household was almost entirely due to a greater number of non-resident dependants, A further interesting inference drawn from this fact would seem to be that a worker who, by social circumstances or otherwise, happens to be charged with the responsibility for a joint household, brings to live with him in the city only those who are 'workers' together with the minimum number of his dependants, leaving the others behind in the village to live on his remittances and other local earnings, if any. How far the reasons for this are social and how far economic it is not possible to state definitely at this stage.

Even more interesting from the social point of view are the conclusions which seem to emerge from a study of the relation which the dependants of the worker in a joint household bear to him. In 185 joint households, the total number of resident dependants came to 360, consisting of 125 males and 235 females. Aoungst the males, 74 or about 60 per cent and amongst the females, 163 or about 70 percent were 'natural' members of the family. The number of non-resident dependants in the same families was 444, consisting of 196 males and 248 females ; and amongst these, 153 or about 78 per cent males and 154 or about 62 per cent females were other than 'natural' members of the family. These percentages (read with the conclusions of the preceding paragraph) strongly point to the inference that the average mill-worker charged with the responsibility of maintaining a joint household ordinarily tries to bring to live with him in the city only his wife and children, and, if additional members of the family have to be brought, those who are 'workers' receive preference ; while the rest are left behind in the village to live on his remittance and other local earnings, if any.

Literacy. Literacy amongst all the persons in the families of the 300 workers comes to 8·3 per cent. The literacy among males comes to 13·8 per cent and amongst females to 1·8 per cent. The literacy amongst workers is 20·1 per cent, amongst resident dependants 6·0 per cent and amongst non-resident dependants 3·2 per cent. All these figures are in accordance with the ideas about literacy commonly held. It is, however, interesting to find that the highest literacy is noticeable between the ages of 15 and 20 years.

Mill-workers expenditure. The following table shows the weights and average monthly expenditure per family of a mill-worker at Cawnpore consisting of 3·1 persons including 1·2 adult males, 0·8 adult female, 0·6 male child under 15 years and 0·5 female child under 15 years.

TABLE 1. EXPENDITURE AND WEIGHTS OF DIFFERENT ITEMS

items	expenditure Rs. As. P.	percentage expenditure food
(1)	(2)	(3)
Wheat	3 1 11	28·7
Rice	1 2 2	10·4
All cereals	5 3 5	48·0
Dal Arhar	0 10 5	6·0
All Pulses	0 14 3	8·2
Ghee	0 14 2	8·1
Oil	0 9 8	5·6
Milk	0 11 3	6·5
Sugar	0 3 1	1·8
Gur	0 2 5	1·4
Meat	0 9 0	5·2
Potato	0 5 6	3·2
Other vegetables	0 5 2	3·0

items	expenditure Rs. As. P.	percentage expenditure	
		food	all items
(1)	2)	(3)	(4)
Spices	0 4 9	2·7	—
Sweets	0 4 11	2·8	—
Fruits	0 2 10	1·6	—
Miscellaneous food	0 3 3	1·9	—
Food Total	10 14 0	100·0	42·0
Clothing	2 0 4	—	7·8
Fuel	1 2 10	—	4·6
Light	0 6 4	—	1·5
Rent	1 11 9	—	6·7
Household requisites	0 11 3	—	2·7
Miscellaneous	8 15 8	—	34·7
All Items	25 14 0	—	100·0

It will be noticed from Table 1 that expenditure on 'miscellaneous' is very high. Since growing expenditure on this head is usually associated with a rising standard of living, it is necessary to analyse the reasons for the comparatively high expenditure on this item. Of the amount of Rs. 8/15/8 spent under this head, Rs. 4/8/2 or more than 50 per cent are spent on remittances, interest and debt repayment, and further Rs. 0/9/2 or 6·4 per cent on litigation. These four items by no means indicate the prosperity or the well-being of the mill-worker in the way expenditure on miscellaneous is usually taken to indicate. If expenditure on these items is taken out and also the expenditure on food items like tobacco, *pan supari* and intoxicants (which also have been placed under miscellaneous), the expenditure on 'miscellaneous' would come down from Rs. 8/15/8 to Rs. 3/4/6 or only 12·7 per cent of the total family expenditure.

An interesting insight into the mill-worker's budget is given by an analysis of his expenditure in relation to his income. The Cawnpore mill-worker spends in a month Rs. 25/14/- and earns in a month Rs. 24/1/5, thus suffering a monthly deficit of Rs. 1/12/7 per family. An examination of the character of this deficit reveals some interesting facts. Actually only 205 families out of 300 under examination had a deficit budget, the deficit being Rs. 3/10/8 per month on an average; the other 95 families had an average monthly surplus of Rs. 2/4/4. Of the 300 families, as many as 179 had to send out money to dependdants living away at the village home, the average remittance per family so sending the money being Rs. 4/4/3. If this remittance is left out of account, the number of families having a deficit budget is reduced from 295 to 137, and of those having a surplus increased from 95 to 163. Remittances are thus responsible for turning 68 surplus budgets into deficit budgets. If remittances be excluded, the average expenditure per family comes down to Rs. 23/5/4, so that an average family would have a surplus of Re. 0/12/1 instead of a deficit of Rs.1 /12/7.

Interest and debt repayment have a similar bearing on the balancing of mill-worker's budget. The combined expenditure on these two items comes to Rs. 2/0/6 p.m. per family. If these items are also left out of account, the number of families having a surplus would rise to 209 and the saving per family having a surplus to Rs. 6/1/10 p.m. In other words, of the 205 families showing a deficit, 114 families suffer a deficit merely on account of their expenditure on the three items of remittance, interest and debt repayments, and but for these three items, the 300 families under examination would show an average saving of Rs. 2/12/7 instead of an average deficit of Rs. 1/12/7 p.m.

Statistics about indebtedness of the mill-worker's family are comparatively less rustworthy than others for obvious reasons. It was found in the course of the inquiry that the worker, while he remembered fairly well the amount he paid to his creditor, did not know in many cases how much of it was interest and how much repayment of the principal. In such cases, of the total amount paid by the worker such portion as was due as interest was considered under the head "interest" and the remaining amount, if any, was credited to 'repayment of debt'.

Of the 300 families, 83·7 per cent were found to be indebted, the average debt being Rs. 136·6 per indebted family. The most important causes of indebtedness were marriage, hartals, payment of rent and revenue at the village home and deficit in monthly budgets. Other causes included expenditure for agricultural purposes at the village home, litigation, illness, unemployment, births, funerals and pilgrimage.

Paper received : 13 April, 1942.

AGRICULTURAL CO-OPERATON IN BENGAL AND REST OF BRITISH INDIA (1918-19 to 1938-39)

By TARINIPROSAD GHOSE and HARISCHANDRA SINHA
Statistical Laboratory, Calcutta.

INTRODUCTORY

To Bengal belongs the credit of initiating the co-operative movement long before it had taken actual shape in the rest of India. The names of two Bengalee pioneers come to mind,—those of Sj. Ambika Charan Ukil and Rai Parbati Shankar Chaudhuri. It is true that the societies started by the former were not co-operative in the strict sense of the term, but there can be perceived in them the rudiments of co-operative principles. The grain banks started by the latter were at first called *Lakshmi Golas*, i.e., granaries presided over by the Hindu Goddess of Plenty, but the name was subsequently changed to *Dharma Golas*, i.e., religious granaries in order not to wound the susceptibilities of non-Hindus. The plan put forward at first was to secure contributions of grain after a bumper harvest and store it up for distribution in a lean year not only among contributors but also among other villagers who were, however, to pay a somewhat higher rate of interest on the amounts borrowed. It is pleasing to recall that the first grain bank started by Rai Parbati Shankar Chaudhuri at Joyganj in the district of Dinajpur as early as in 1892 continued its useful career as an unregistered society till December, 1914, after which it was registered in the usual way. Even if we leave this out and consider only co-operative societies in the proper sense of the term, the earliest institution was started at the village Kushmore (P.O. Labpur in the district of Birbhum) on July 12, 1902. About forty years have elapsed since then and it is time a study was made about the progress and development of the co-operative movement. It is true that there have been quite a number of such periodical reviews at the series of Conferences of Registrars of Co-operative Societies of the different provinces of British India and of some of the Indian States, but these were mainly concerned with administrative problems. There have been also deeper and more incisive studies, e.g. *Maclagan Committee on Co-operation in India* (1915) and Prof. J. P. Niyogi's *The Co-operative Movement in Bengal*. But it is felt by the authors that there is scope for a fresh study on statistical lines. In particular, the authors desire to compare the standard reached in Bengal with that in the rest of British India, not only on broad lines with regard to the movement in general, but also in somewhat greater detail in the case of agricultural societies, which form the backbone of the system, in order that some light may be thrown on the obscurities hiding the strength and weakness of such societies in Bengal as compared with the rest of British India. It is hoped, however, to undertake similar studies in future with regard to other components of the system such as non-agricultural co-operative societies, central banks etc.

INDIAN STATES OMITTED

2. It will be noticed that Indian States have been omitted throughout, and that for two reasons. Firstly because different States began to report at different times and secondly because different standards have been reached by different States. For instance, during the year 1938-39, the number of societies per 1,00,000 inhabitants was 105·1 in Gwalior and only 22·3 in Hyderabad, although the number of societies was nearly the same viz., 3,992 and 3,610 respectively. It is true that wide variations are observed also in some of the provinces of British India such as Coorg and Ajmer Merwar which record very high *per capita* figures for number of societies, but the former had only 300 societies and the later only 721 societies out of a total of 1,04,307 societies for the whole of British India during 1938-39.

DOES BRITISH INDIA AFFORD A SATISFACTORY STANDARD ?

3. We must confess that even if we leave out Indian States and consider only British India, we do not have a fully satisfactory standard for the purpose of comparison. In the first place, there are wide differences in the progress of co-operation even among the major provinces with not less than 10 million population, as compiled in Table 1 from *Statistical Statements Relating to the Co-operative Movement in India during the year* 1938-39.

4. In the second place, all provinces are not equally productive, as will appear from Table 1 col. 5 giving the money value of principal crops in 1933-34, compiled from the *Review of the Trade of India* in

1934-35. This disparity is not, however, due to varying population in the different provinces, a fact apparent from the last column of the Table, compiled from the *Statistical Statements Relating to the Co-operative Movement in India during the year* 1933-34.

TABLE 1. PROGRESS OF CO-OPERATION AND VALUE OF CROPS AND POPULATION IN DIFFERENT PROVINCES

province	1938-39			1933-34	
	societies per 1,00,000 inhabitants	members of primary societies per 1,000 inhabitants	working capital per head of population (Rs.)	value of principal crops (Rs. Lakhs)	population (Millions)
(1)	(2)	(3)	(4)	(5)	(6)
Madras	27·6	20·9	4·6	83·17	46·7
Bombay	25·9	30·0	8·2	60·52	21·9
Bengal	57·8	17·7	3·8	95·56	50·1
Behar	20·7	5·7	1·3	55·30*	37·7*
U. P.	21·1	10·2	0·6	85·70	48·4
Punjab	92·6	35·1	6·5	40·10	23·6
C.P. & Berar	28·5	5·6	3·2	30·33	15·5

* Figures for Bihar & Orissa.

5. Thus it is clear that even British India does not afford a satisfactory standard. An alternative procedure is to compare Bengal with other provinces in succession and to see to what extent she is in a better or in a worse position. But this would be ruled out, because the results would be vitiated by the difference in the economic levels of the respective provinces. We have therefore thought it advisable to compare Bengal with the rest of British India as a whole, composed as it is of productive as well as unproductive provinces. It must be remembered, however, that as a result of such comparison Bengal would tend to appear in too favourable a light, for the all-India standard must necessarily be dragged down by large economically backward areas. This fact should always be borne in mind in interpreting the results of the present statistical analysis. In some cases, the corresponding figures for the two Presidencies other than Bengal, viz., Bombay and Madras have been added in order to facilitate mutual comparison. It will be seen that we have not considered the Punjab, because she is, on all hands, co-operatively the most progressive area.

SOURCES OF DATA

6. The data are all computed from the official publication *Statements showing progress of the Co-operative Movement in India*, re-styled *Statistical Statements Relating to the Co-operative Movement, in India* since 1927-28. The statistics were considerably modified following the recommendation of the *Maclagan Committee on Co-operation in* 1915. It has therefore been thought advisable to omit the previous years for which the figures are not on the same basis as subsequently. Again, although individual members of Central and Provincial Banks, Banking Unions and Guaranteeing Unions were being included in the total of figures, they have been omitted since 1917-18, in order to avoid double counting, for it is assumed that these have already figured once as members of an agricultural or non-agricultural primary society. For these and other reasons, the data have been considered from 1918-19 onwards till 1938-39, the latest year for which the figures are available at present. With effect from April 1, 1937, Burma has been separated from British India and in order to maintain continuity, Burma figures have been omitted throughout. Thus the figures for Br. India from 1918-19 to 1936-37 quoted in the present paper refer to British India proper excluding Burma, and will not therefore agree with the corresponding figures published in the official *Statistical Statements*, although there will be agreement from 1937-38 onwards.

ANOMALIES IN DATA

7. Inspite of these expedients, all anomalies in the data could not be eliminated. To start with, different provinces make up their annual reports on co-operation on different dates: Assam: 31st March; Ajmer-Merwar, Bengal, Bombay, Central Provinces and Berar, Coorg, Madras, Orissa, United Provinces: 30th June; Hyderabad Administered Area, 7th July: Delhi, Punjab, North-West Frontier Province: 31st July; Bihar: 31st December.

8. It is however clear that the majority of the provinces, whether numerically, or from the point of view of number of societies, or of membership, or of total working capital make up their reports as on

30th June, on which the general agricultural year of India ends and which has therefore been taken as the closing date in the present paper. It is hoped, however, that this will not affect the final results, which are based on year-to-year comparisons and not on individual annual figures.

9. Secondly, with effect from 1937-38 particulars relating to land mortgage banks have been shown separately in a new subsidiary Statement B-1, and not included as before in the General Statements A, B and C. Unfortunately, the figures for Bengal are published in such a form that they cannot be incorporated into the General Statements A, B and C, so as to maintain continuity with the period prior to 1937-38. Even on reference to the Registrar of Co-operative Societies, Bengal, the desired information could not be obtained, for the principle of making up the table had been changed. We were therefore reluctantly compelled to omit the details of the working capital of agricultural co-operative societies for 1937-38 and 1938-39. But to maintain uniformity we found it necessary to add the figures relating to the number of societies, the number of members and the total working capital for 1937-38 and 1938-39 for Land Mortgage Banks to those for agricultural societies. There are two misprints in the *Statistical Statements* for 1926-27 and 1931-32, which have been corrected by us. "Other Funds" have been shown separately since 1934-35 in the *Statistical Statesments*. But in order to maintain continuity, they have been included in the "Reserve Fund" throughout.

CORRECTED DATA

10. The finally adjusted figures as compiled from the *Statistical Statements* are given in Table 2 for Bengal and Table 3 for the rest of British India. In the last column of each of the tables, the average of the monthly index number of wholesale prices in Calcutta from July to June each year has been added. It will be seen that during the period under review there was considerable fluctuation in the price level, on account of the post-war boom of 1919-20 and the crash thereafter, the Great Depression of 1929-30 and the subsequent recovery and finally the boom at the out break of the present war. The figures for working capital and of their main constituents have therefore to be interpreted in the light of this variation in the price level, for which the index numbers have been inserted.

GENERAL FEATURES

11. The figures quoted above, whether for agricultural societies or for non-agricultural societies, indicate almost uninterrupted progress. But if they are properly scrutinised, temporary declines will be noticed in the case of agricultural societies, although not in the case of non-agricultural societies. Thus on reference to column (2) of Table 2, it will be seen that the number of agricultural societies in Bengal began to decline from 1932-33, which was arrested as late as in 1936-37. If the corresponding series for the rest of British India, as quoted in column (2) of Table 3, are looked into, it will be found that there was a set-back in 1931-32 followed by a recovery with effect from 1934-35. With regard to the number of members of agricultural societies the decline lasted from 1931-32 to 1937-38 in Bengal and from 1931-32 to 1933-34 in the rest of British India, as will appear from column (4) in Tables 2 and 3 each.

DEFLATED WORKING CAPITAL

12. Such temporary set-backs, however, are not found in the case of working capital either for Bengal or for the rest of British India. But if allowance is made for variation in the price level, by dividing the value figure for each year by the corresponding index number, and multiplying the quotient by 100 so as to get the appropriate deflated value, the following results are obtained as given in Table 4. On reference to columns (2), (3), (6) and (7) it will be seen that there were temporary set-backs and revivals from 1933-34 onwards both in Bengal and outside in the case of agricultural societies. The considerable increase recorded for 1938-39 for Madras (column 7, last line) is accounted for by the inclusion of land mortgage banks. The working capital of Madras Land Mortgage Banks rose from Rs. 156 lakhs in 1937-38 to Rs. 412 lakhs in 1938-39. If allowance is made for variation in the price level, the increase was from nearly Rs. 161 lakhs to Rs. 425 lakhs. In the case of non-agricultural societies, however, there is un-interrupted progress throughout with the exception of Bombay for 1935-36. But the rates of increase are not identical.

13. From the above it is clear that although the figures for Bengal are broadly similar to those for the rest of British India, there are certain points of difference. A more intensive study on statistical lines is therefore called for in order to scrutinise to what extent the figures for Bengal differ from those elsewhere. It is a statistical commonplace to say that the respective rates of progress may be conveniently studied with the help of the relative trends. In the case of cumulating data of the nature of co-operative statistics

TABLE 2. STATISTICS OF CO-OPERATION IN BENGAL

year (July to June)	number of societies (in thousands)		number of members (in thousands)		working capital (in lakhs of rupees)		components of working capital of agricultural societies (in lakhs of rupees)							Calcutta index number of whole-sale prices	year (July to June)
							capital & reserves		loans & deposits						
	agricul-tural	non-agricul-tural	agricul-tural societies	non-agricul-tural societies	agricul-tural societies	non-agricul-tural societies	share capital	reserve fund	from members	from non-members	from societies	from prov & central banks	from govt.		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
1918—19	3·92	0·26	134·76	36·09	92·35	36·21	1·28	14·59	5·68	4·53	0·98	65·05	0·25	186	1918—19
—20	4·98	0·36	157·16	50·68	111·12	47·86	2·30	16·67	6·14	5·08	0·59	80·11	0·21	205	—20
—21	5·85	0·44	171·83	60·17	128·26	63·38	3·82	17·19	6·59	6·06	0·93	93·54	0·13	188	—21
—22	6·15	0·45	180·78	68·72	138·78	73·93	5·46	20·05	6·91	6·87	0·95	98·27	0·27	180	—22
—23	7·21	0·51	201·24	79·72	156·74	90·91	8·02	23·58	7·46	8·13	1·08	108·25	0·23	173	—23
—24	8·64	0·60	234·82	91·91	188·54	105·88	11·56	27·74	8·39	10·05	1·20	129·32	0·27	170	—24
—25	10·20	0·75	268·71	101·86	222·79	125·51	16·34	33·04	10·08	11·42	0·82	150·68	0·41	168	—25
—26	11·64	1·03	316·76	129·83	267·24	156·34	23·09	38·19	12·09	12·14	1·28	180·04	0·41	154	—26
—27	13·95	1·33	379·24	139·86	338·34	177·04	31·32	46·72	13·76	13·72	0·85	231·55	0·41	146	—27
—28	16·43	1·52	441·21	160·63	408·25	207·83	39·35	57·24	15·77	15·07	0·83	279·53	0·46	147	—28
—29	17·97	1·73	476·64	192·98	470·15	243·43	40·80	69·68	17·65	16·83	1·04	315·67	0·48	143	—29
—30	20·44	1·91	512·43	193·71	552·48	276·88	58·56	85·44	19·71	16·61	0·88	370·83	0·45	132	—30
—31	21·46	2·03	535·29	224·90	561·23	312·93	58·71	103·06	23·37	16·37	1·12	358·18	0·42	103	—31
—32	21·49	2·12	529·44	239·45	578·55	338·80	60·69	122·24	21·46	18·62	1·14	354·01	0·38	94	—32
—33	21·34	2·18	523·44	247·71	592·03	376·76	61·91	140·00	21·34	18·83	1·08	348·48	0·39	88	—33
—34	21·22	2·20	512·82	262·27	593·35	425·36	62·15	155·35	21·21	18·80	1·10	334·36	0·37	89	—34
—35	21·12	2·19	512·95	267·87	609·11	467·25	62·39	173·00	20·19	16·62	2·35	334·21	0·36	89	—35
—36	21·11	2·28	506·61	286·46	615·21	502·50	61·31	185·10	21·28	15·76	2·44	328·97	0·36	91	—36
—37	21·26	2·83	507·31	303·71	618·43	540·29	59·97	104·82	20·69	15·90	2·45	324·24	0·36	96	—37
—38	21·26	2·86	498·44	321·50	616·44	574·25								100	—38
—39	27·55	2·99	599·48	332·83	626·97	603·93								97	—39

TABLE 3. STATISTICS OF CO-OPERATION IN THE REST OF BRITISH INDIA : EXCLUDING BENGAL

(year (July to June)	number of societies (in thousands)		number of members (in thousands)		working capital (in lakhs of rupees)		components of working capital of agricultural societies (in lakhs of rupees): capital & reserve		loans & deposits					Calcutta index number of whole-sale prices	year (July to June)
	agricul-tural	non-agricul-tural	agricul-tural societies	non-agricul-tural societies	agricul-tural societies	non-agricul-tural societies	share capital	reserve fund	from members	from non-members	from societies	from prov & central banks	from govt.		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
1918—19	19·97	1·30	691·90	172·29	541·93	132·29	77·85	74·72	38·57	40·03	5·91	301·98	2·86	186	1918—19
—20	24·81	1·75	845·50	226·44	654·20	177·06	85·32	87·81	47·94	46·93	5·01	378·70	2·48	205	—20
—21	28·87	2·23	986·22	255·97	804·84	219·16	93·16	102·09	56·70	56·43	6·15	480·73	9·57	188	—21
—22	32·29	2·53	1127·78	301·88	922·23	288·31	101·44	117·96	64·61	58·77	7·37	561·73	10·36	180	—22
—23	34·51	2·72	1192·15	327·23	1028·91	357·98	106·07	134·16	70·75	62·86	4·93	640·29	9·84	173	—23
—24	37·37	3·12	1320·12	361·03	1157·43	417·46	115·59	155·02	84·62	66·17	5·48	722·67	7·88	170	—24
—25	41·52	3·73	1474·15	407·12	1309·80	490·13	137·97	179·83	92·11	70·24	8·08	872·02	9·55	168	—25
—26	45·03	4·71	1608·03	478·74	1610·33	585·86	160·13	208·88	110·32	76·03	9·16	1027·10	18·71	154	—26
—27	49·26	5·36	1793·59	530·87	1898·74	683·03	187·91	241·46	114·22	87·52	6·51	1246·68	15·00	146	—27
—28	51·85	5·93	1949·90	592·78	2107·41	793·09	215·94	280·75	125·78	95·85	12·58	1360·20	16·30	147	—28
—29	54·05	6·27	2036·21	631·63	2292·56	920·43	238·76	321·58	129·89	109·91	19·93	1156·50	15·99	143	—29
—30	55·42	6·43	2102·76	684·76	2478·46	1012·08	260·14	369·93	133·89	112·06	19·61	1567·63	15·22	132	—30
—31	55·81	6·50	2121·05	713·27	2552·41	1078·30	274·58	414·02	134·11	118·76	20·15	1574·90	15·89	103	—31
—32	55·56	6·56	2068·94	730·12	2444·98	1121·82	276·18	449·62	129·32	116·41	18·99	1441·00	13·47	94	—32
—33	55·06	6·65	2015·74	759·28	2367·75	1202·25	277·93	489·93	123·79	116·64	18·86	1328·12	12·49	88	—33
—34	54·92	6·78	1984·40	830·85	2334·38	1274·82	276·13	540·07	120·41	106·94	19·83	1251·37	9·63	89	—34
—35	55·67	7·07	2005·53	868·58	2340·86	1378·51	274·22	579·09	116·62	116·06	19·18	1226·32	9·37	89	—35
—36	56·98	7·63	2054·19	898·07	2385·47	1499·61	282·25	610·51	106·53	111·50	17·73	1249·69	7·62	91	—36
—37	58·75	8·31	2164·22	967·04	2399·09	1598·77	285·02	627·58	102·59	111·06	17·06	1249·40	6·39	96	—37
—38	60·17	8·95	2277·83	1041·07	2623·74	1680·43								100	—38
—39	63·16	9·62	2576·20	1098·60	2686·53	1786·34								97	—39

TABLE 4. DEFLATED VALUE OF WORKING CAPITAL (in lakhs of rupees)

Year	agricultural society		non-agricultural society		agricultural society		non-agricultural society	
	Bengal	rest. of Br. India	Bengal	rest. of Br. India	Bombay	Madras	Bombay	Madras
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1918—19	49 ·65	291 ·36	19 ·47	71 ·12	44 ·95	66 ·46	33 ·49	25 ·96
—20	54 ·21	319 ·12	24 ·35	86 ·37	53 ·85	77 ·89	41 ·16	31 ·08
—21	68 ·23	428 ·10	33 ·71	116 ·57	72 ·24	99 ·80	58 ·27	38 ·56
—22	77 ·10	512 ·35	41 ·07	160 ·17	88 ·96	126 ·21	82 ·73	48 ·78
—23	90 ·60	594 ·74	52 ·55	206 ·92	101 ·73	156 ·39	108 ·58	62 ·29
—24	110 ·91	680 ·84	62 ·28	245 ·57	116 ·25	188 ·74	117 ·00	77 ·34
—25	132 ·61	815 ·36	74 ·71	291 ·75	148 ·80	217 ·70	134 ·31	98 ·55
—26	173 ·53	1045 ·67	101 ·52	380 ·43	204 ·39	264 ·58	172 ·07	125 ·57
—27	231 ·74	1300 ·51	121 ·26	467 ·83	248 ·97	343 ·15	204 ·33	155 ·20
—28	277 ·72	1433 ·61	141 ·38	539 ·52	252 ·12	398 ·48	236 ·23	174 ·04
—29	328 ·78	1603 ·19	170 ·23	643 ·66	258 ·04	439 ·99	271 ·50	209 ·15
—30	418 ·54	1877 ·62	209 ·76	766 ·73	310 ·66	512 ·96	330 ·99	238 ·25
—31	544 ·88	2478 ·07	303 ·82	1046 ·89	436 ·22	649 ·62	472 ·36	321 ·70
—32	615 ·48	2601 ·04	360 ·43	1193 ·43	469 ·12	630 ·64	552 ·72	349 ·58
—33	672 ·76	2690 ·63	428 ·13	1366 ·19	481 ·78	619 ·78	646 ·26	390 ·22
—34	666 ·68	2622 ·90	477 ·94	1432 ·38	462 ·40	584 ·22	698 ·08	391 ·80
—35	684 ·40	2630 ·18	525 ·00	1548 ·89	454 ·43	614 ·73	764 ·52	415 ·84
—36	676 ·06	2621 ·40	552 ·20	1647 ·92	357 ·50	642 ·31	678 ·32	430 ·87
—37	644 ·20	2499 ·05	562 ·80	1665 ·39	351 ·44	633 ·41	679 ·49	433 ·07
—38	616 ·44	2623 ·74	574 ·25	1680 ·43	362 ·89	671 ·32	688 ·89	444 ·43
—39	646 ·37	2769 ·62	622 ·61	1841 ·59	377 ·36	1013 ·40	756 ·17	488 ·82

quoted in Tables 2 and 3, a Gompertz or Logistic curve immediately suggests itself for studying the law of growth. As Mr. R. B. Prescott[6] points out, such a curve is quite appropriate when there are an initial period of experimentation, a subsequent period of growth into social fabric, a third stage during which this rate of growth is successively accelerated, becomes stationary and retarded, and final stage when there is no further growth,—in other words, a period of economic equilibrium, provided of course further expansion in population, income etc., is stopped. So far as our present data are concerned, we may tentatively conclude that they are in the second and third stages, outgrowing the early experimental years on the one hand and not reaching the final equilibrium period on the other. How far this is true will appear from the Table 5 below, which gives details about the progress of co-operation in relation to population during the first

TABLE 5. PROGRESS OF CO-OPERATION IN RELATION TO POPULATION

area	1918—1919				1938—1939.			
	no. of societies per1,00,000 inhabitants	no. of members of primary societies per 1,000 inhabitants	amount of working capital per head of population (Rs.)	deflated value of working capital per head of population (Rs.)	no. of societies per 1,00,000 inhabitants	no. of members of primary societies per 1,000 inhabitants	amount of working capital per head of population (Rs.)	deflated value of working capital per head of population (Rs.)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Br. India Ex-Bengal	13 ·2	4 ·9	0 ·7	0 ·4	32 ·9	16 ·5	3 ·4	3 ·5
Bengal	9 ·5	3 ·8	0 ·5	0 ·2	57 ·8	17 ·7	3 ·7	3 ·9
Bombay	11 ·2	9 ·8	1 ·1	0 ·6	25 ·9	30 ·0	8 ·2	8 ·4
Madras	8 ·9	5 ·9	0 ·8	0 ·4	27 ·6	20 ·9	4 ·6	4 ·8

and the last years of the period under investigation, in Bengal and elsewhere. On reference to columns (2) and (6), it will be seen that the number of societies per 1,00,000 inhabitants have risen from 13.2 in 1918-19 to 32 ·9 in 1938-39 in British India exclusive of Bengal. The number of members of primary societies per 1,000 inhabitants has increased from 4 ·9 to 16 ·5 as shown in columns (3) and (7). The position in different Presidencies is not, however, identical. In Bengal in particular the number of societies has risen more steeply than elsewhere. In any case, it is quite clear that we are yet at a great distance from final "saturation point", either in Bengal or elsewhere although some progress has undoubtedly been achieved.

TRENDS FOR BENGAL AND REST OF BRITISH INDIA

14. We thus come to the conclusion that the trends for the period under review may be regarded as approximately linear, provided we treat all data logarithmically,—for number of societies, for number of members, and for working capital, both in the case of agricultural and non-agricultural societies, as given in columns (2) to (7) in Table 2 for Bengal and in Table 3 for the rest of British India, all the required deflated values for working capital being given in Table 4. The results obtained are given in Table 6, y, being the smoothed value of the variable concerned in each case, and t being the time measured in years from the middle of the series viz., 1928-29. Thus the annual rate of growth of the number of agricultural societies in Bengal works out at 0·039 whereas the corresponding figure for the rest of British India is found to be 0·020. It may be argued that this observed difference is not significant in the sense that such variation may easily arise from sampling fluctuations. To test whether this is so or not we apply Fisher's t-test, the results being given in Table 7. It will be noted that barring the cases of working capital of non-agricultural societies, both ordinary and deflated, all the other rates of growth are found to be significantly higher in Bengal.

TABLE 6. RATES OF PROGRESS OF CO-OPERATION IN BENGAL AND THE REST OF BRITISH INDIA (1918-19 to 1938-39)

no.	trends for	Bengal	rest of British India	value of 't'
1	No of agricultural Societies.	Log $y=4\cdot128795+0\cdot038538\,t$	log $y=4\cdot654871+0\cdot020315\,t$	4·42*
2	No of non-agricultural Societies.	Log $y=3\cdot089000+0\cdot051829\,t$	log $y=3\cdot679052+0\cdot037419\,t$	3·00*
3	No of members of agricultural Societies.	Log $y=5\cdot547652+0\cdot031370\,t$	log $y=6\cdot215390+0\cdot022632\,t$	2·12*
4	No of members of non-agricultural Societies.	Log $y=5\cdot177748+0\cdot045579\,t$	log $y=5\cdot733976+0\cdot036589\,t$	2·71*
5	Total working capital of agricultural Societies.	Log $y=7\cdot527205+0\cdot044553\,t$	log $y=8\cdot224243+0\cdot031887\,t$	2·44*
6	Total working capital of non-agricultural Societies.	Log $y=7\cdot307319+0\cdot059775\,t$	log $y=7\cdot846210+0\cdot053068\,t$	1·73
7	Deflated working Capital of agricultural Societies.	Log $y=5\cdot414971+0\cdot064417\,t$	log $y=6\cdot112010+0\cdot051751\,t$	2·08*
8	Deflated working Capital of Non-agricultural Societies.	Log $y=4\cdot889171+0\cdot073912\,t$	log $y=5\cdot733976+0\cdot072932\,t$	0·08

Degrees of freedom 38 ; value of "t" at five per cent level = 2·03. Results significant at five per cent level are marked with asterisks. The tentative conclusion for such significant values is that the rate of growth for Bengal is higher than that for the rest of British India.

NUMBER OF MEMBERS PER SOCIETY

15. The question now is, has the increase in membership kept pace with the increase in the number of societies, both agricultural and non-agricultural? What is the position in Bengal compared with the rest of British India ? The answers to these questions are to be found in Table 7. Column (2) giving the number of members per agricultural society in Bengal is obtained on dividing column (4) by column (2) of Table 2. Column (3) giving the corresponding figure for non-agricultural soicieties is similarly obtained on dividing column (5) by column (3) of Table 3. Columns (4) and (5) are similarly calculated from the corresponding figures for the rest of British India in Table 3. It will be seen that in Bengal the number of members per society has gone down both in the case of agricultural and non-agricultural societies, whereas in the rest of British India, only the number of members per non-agricultural society has decreased but the corresponding figure for agricultural societies has steadily gone up. Low membership means smaller societies for which the overhead expenses must necessarily be higher. There is therefore a possibility that in the case of agricultural societies, quality has been sacrificed for quantity in the case of Bengal as compared with the rest of British India. But in the case of non-agricultural societies the position in Bengal seems to be not unsatisfactory as compared with the rest of British India. In order to judge the extent of the disparity between the two more precisely, columns (6) and (7) have been inserted in Table 7, in which the figure for Bengal is expressed as a percentage of the corresponding figure for the rest of British India. Thus if the

figures in columns (6) and (7) are below 100, Bengal's position is worse than that of the rest of British India, whereas if the figure is above 100, Bengal's position is better to that extent. Column (6) shows that from the point of view of membership, the condition of agricultural societies has steadily deteriorated as compared with the rest of British India but with regard to non-agricultural societies the position is not so bad, as will appear from a study of column (7), although the relatively favourable position of 1923-24 has all but passed away. The position of Bengal vis-a-vis that of Bombay and Madras is shown in columns (2), (3) and (8) to (11). On a study of columns (2), (8) and (10), it will be seen that agricultural societies have much fewer members in Bengal than in Bombay or Madras. The same conclusion is also apparent in the case of non-agricultural societies from columns (3), (9) and (11). It will also be seen that the extent of the superiority over the British India standard in the case of Bengal is being consistently damped down, whereas in the cases of Bombay and Madras, this superiority is being enhanced.

Table 7. Number of members per society

year	Bengal		Rest of British India		disparity index between the two		Bombay		Madras	
	agri-cultural	non-agricul-tural	agri-cultural	non-agricul-tural	agri-cultural 100(2)/(4)	non-agricul-tural 100(3)/(5)	agri-cultural	non-agricul-tural	agri-cultural	non-agricul-tural
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1918—19	34·3	137·8	34·7	132·7	99	104	71·3	212·4	55·6	152·7
—20	31·6	140·8	34·1	129·6	93	109	74·6	191·1	54·8	150·3
—21	29·4	137·4	34·2	114·7	86	120	75·3	177·4	53·4	143·2
—22	29·4	151·7	34·9	119·1	84	127	75·6	199·9	53·0	145·8
—23	27·9	155·4	34·5	120·1	81	129	72·3	211·2	53·7	146·0
—24	27·2	152·4	35·3	115·6	77	132	72·3	212·8	54·7	136·8
—25	26·3	136·4	35·5	109·2	74	125	71·9	215·9	53·8	135·8
—26	27·2	125·8	35·7	101·6	76	124	70·1	248·1	53·9	141·2
—27	27·2	104·9	36·4	99·1	75	106	70·0	249·1	53·5	143·2
—28	26·9	105·7	37·6	99·9	72	106	75·8	248·2	53·3	146·3
—29	26·5	111·4	37·7	100·8	71	111	74·1	255·3	53·4	148·3
—30	25·1	102·2	37·9	106·6	66	96	71·9	263·7	54·5	157·5
—31	24·9	111·0	38·0	109·7	66	101	69·4	265·8	55·7	169·3
—32	24·6	113·8	37·2	111·3	66	102	66·6	272·7	53·8	174·4
—33	24·5	113·4	36·6	114·1	67	99	65·7	285·8	52·4	185·4
—34	24·2	119·2	36·1	118·5	67	101	65·2	299·9	51·0	196·7
—35	24·3	122·5	36·0	122·8	68	100	63·6	346·4	50·3	199·8
—36	24·0	125·8	36·1	117·7	67	107	67·1	331·1	51·0	202·1
—37	23·9	107·4	36·8	116·4	65	92	67·3	331·8	52·2	200·0
—38	23·4	112·4	37·9	116·3	62	97	65·6	352·1	54·8	193·4
—39	21·8	111·4	40·8	114·3	53	98	64·8	345·9	58·6	192·0

WORKING CAPITAL PER SOCIETY

16. Let us now consider the matter not from the point of view of membership but of working capital per society. The necessary figures are given in Table 8 compiled in the same way as Table 7. On a scrutiny of columns (6) and (7), we come to the same conclusion as before viz., that Bengal agricultural societies have fared worse than those in the rest of British India, but in the case of non-agricultural societies, the standard is about the same throughout British India. The hopelessly inferior position of agricultural societies in Bengal as compared with that in Bombay and Madras can be easily seen from columns (2), (8), and (10). On a study of columns (3) and (9) it will be seen that even in the case of non-agricultural societies. Bengal has much poorer working capital than Bombay. On reference to column (11) it will appear that even so far as Madras is concerned, although Bengal began with greater working capital per society in 1918-19, she had consistently less working capital since 1926-27.

WORKING CAPITAL PER MEMBER

17. It may be argued that in the case of Bengal agricultural societies, both membership and working capital have gone down as compared with the rest of British India, but it does not also mean that working

TABLE 8. WORKING CAPITAL PER SOCIETY (In thousands of rupees)

year	Bengal		rest of Br. India		disparity index between the two		Bombay		Madras	
	agri-cultural society	non-agricul-tural society	agri-cultural society Rs.	non-agricul-tural society Rs.	agri-cultural society 100(2)/(4)	non-agricul-tural society 100(3)/(5)	agri-cultural society	non-agricul-tural society	agri-cultural society	non-agricu-turall society
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1918—19	2 ·35	13 ·82	2 ·71	10 ·19	87	136	4 ·81	22 ·25	4 ·01	10 ·38
—20	2 ·23	13 ·29	2 ·64	10 ·14	85	131	5 ·28	21 ·20	3 ·79	9 ·94
—21	2 ·19	14 ·47	2 ·79	9 ·82	79	147	5 ·70	22 ·59	3 ·56	9 ·03
—22	2 ·26	16 ·32	2 ·86	11 ·38	79	143	5 ·82	26 ·54	3 ·61	10 ·26
—23	2 ·17	17 ·12	2 ·98	13 ·14	73	130	6 ·13	33 ·01	3 ·76	11 ·16
—24	2 ·18	17 ·56	3 ·10	13 ·37	71	131	6 ·48	33 ·15	3 ·86	11 ·28
—25	2 ·18	16 ·80	3 ·30	13 ·15	66	128	7 ·40	34 ·93	3 ·86	12 ·62
—26	2 ·30	15 ·15	3 ·56	12 ·43	65	122	8 ·14	39 ·49	4 ·00	13 ·85
—27	2 ·42	13 ·33	3 ·85	12 ·76	63	105	8 ·48	43 ·61	4 ·38	15 ·10
—28	2 ·49	13 ·67	4 ·06	13 ·37	61	102	8 ·36	45 ·16	4 ·70	16 ·06
—29	2 ·62	14 ·05	4 ·24	14 ·69	62	96	8 ·08	50 ·03	4 ·86	18 ·03
—30	2 ·70	14 ·48	4 ·47	15 ·75	60	92	8 ·58	52 ·58	5 ·17	19 ·07
—31	2 ·61	15 ·45	4 ·57	16 ·58	57	93	9 ·19	55 ·48	5 ·22	21 ·17
—32	2 ·69	16 ·10	4 ·40	17 ·10	61	94	9 ·11	57 ·86	4 ·72	21 ·75
—33	2 ·77	17 ·25	4 ·30	18 ·07	65	96	8 ·80	61 ·82	4 ·55	23 ·90
—34	2 ·80	19 ·33	4 ·25	18 ·80	66	103	8 ·65	67 ·53	4 ·34	24 ·37
—35	2 ·88	21 ·37	4 ·21	19 ·49	69	110	8 ·55	75 ·10	4 ·71	25 ·83
—36	2 ·91	22 ·06	4 ·19	19 ·65	70	112	8 ·32	75 ·00	5 ·09	26 ·75
—37	2 ·91	19 ·10	4 ·08	19 ·25	71	99	8 ·36	73 ·21	5 ·49	28 ·00
—38	2 ·90	20 ·07	4 ·36	18 ·77	67	107	9 ·00	75 ·21	6 ·04	27 ·32
—39	2 ·28	20 ·21	4 ·25	18 ·58	54	109	9 ·04	78 ·36	8 ·55	26 ·77

capital per member is worse in Bengal than outside as can be seen from Table 9. But it should be noted that here, as elsewhere, the position in Bengal is definitely inferior to that in Bombay and Madras, particularly for agricultural societies.

TABLE 9. WORKING CAPITAL PER MEMBER

year (July to June)	Bengal		rest of British India		disparity index between the two		Bombay		Madras	
	agri-cultural society Rs.	non-agricul-tural society Rs.	agri-cultural society Rs.	non-agricul-tural society Rs.	agri-cultural society	non-agricul-tural society	agri-cultural society Rs.	non-agricul-tural society Rs.	agri cultural society Rs.	non-agricul-tural society Rs.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1918-19	68 ·6	100 ·3	78 ·2	76 ·8	88	131	67 ·5	104 ·7	72 ·2	68 ·0
—20	70 ·7	94 ·4	77 ·3	78 ·2	92	121	70 ·8	110 ·9	69 ·1	66 ·1
—21	74 ·6	105 ·3	81 ·5	85 ·6	92	123	75 ·7	127 ·3	66 ·9	63 ·1
—22	76 ·8	107 ·6	81 ·8	95 ·5	94	123	77 ·0	132 ·8	68 ·2	70 ·4
—23	77 ·9	110 ·2	86 ·4	109 ·4	90	101	84 ·9	156 ·3	70 ·0	76 ·4
—24	80 ·2	115 ·2	87 ·7	115 ·7	91	100	89 ·6	155 ·8	70 ·7	82 ·4
—25	83 ·0	123 ·2	93 ·0	120 ·4	89	102	103 ·0	161 ·8	71 ·8	92 ·9
—26	84 ·4	120 ·4	99 ·7	122 ·3	85	99	116 ·0	159 ·2	74 ·3	98 ·1
—27	89 ·1	127 ·1	105 ·9	128 ·7	84	99	121 ·1	175 ·1	81 ·8	105 ·4
—28	92 ·4	129 ·4	108 ·1	133 ·8	86	97	110 ·2	181 ·9	88 ·2	109 ·8
—29	98 ·7	126 ·1	112 ·5	145 ·7	88	87	109 ·0	196 ·0	91 ·1	121 ·3
—30	107 ·7	141 ·7	118 ·0	147 ·8	91	96	119 ·3	199 ·4	94 ·9	121 ·1
—31	105 ·0	139 ·2	120 ·4	151 ·2	81	92	132 ·5	208 ·7	93 ·6	125 ·0
—32	109 ·4	141 ·4	118 ·3	153 ·6	93	92	136 ·8	212 ·1	87 ·8	124 ·7
—33	113 ·1	152 ·1	117 ·5	158 ·4	96	96	133 ·9	216 ·3	86 ·7	128 ·9
—34	115 ·6	162 ·1	117 ·7	158 ·6	98	102	132 ·8	200 ·5	87 ·0	123 ·9
—35	118 ·7	174 ·5	116 ·8	158 ·7	102	110	134 ·5	216 ·8	93 ·7	129 ·2
—36	121 ·4	175 ·3	116 ·0	167 ·0	105	105	124 ·0	226 ·5	99 ·8	132 ·3
—37	121 ·7	177 ·9	111 ·0	165 ·3	110	108	124 ·1	220 ·7	105 ·3	140 ·0
—38	123 ·7	178 ·6	115 ·2	161 ·4	107	111	137 ·2	213 ·6	110 ·3	141 ·2
—39	104 ·6	181 ·5	104 ·3	162 ·5	100	112	139 ·6	226 ·5	145 ·7	139 ·5

WORKING CAPITAL OF AGRICULTURAL SOCIETIES

18. That some difference exists in the case of Bengal agricultural societies is an undoubted fact. What is the reason for this ? Some light may be thrown on the matter by a proper scrutiny of the different elements of the working capital viz., (a) Share capital, (b) Reserve Fund, (c) Members' loans and deposits, (d) Non-members' loans and deposits, (e) Societies', Provincial and Central Banks' loans and deposits, and (f) Government loans and deposits as shown in columns (8) to (14) in Table 2 for Bengal and in Table 3 for the rest of British India. Most of these series record increases but their respective rates cannot be easily compared. Moreover they are all affected by the fluctuation in the price level shown in column (15) of each of the Tables 2 and 3. To overcome these difficulties, each of the above components has been expressed as percentages of the total working capital in Table 10. Thus column (2) of this table for percentage of share capital to total working capital in Bengal is obtained by expressing column (8) in Table 2 as a percentage of column (6) in the same table. Column (3) of Table 10 is similarly obtained from the corresponding figures in Table 3 for the rest of British India. The disparity index showing the relationship between Bengal and the rest of British India is obtained on dividing column (2) for the former by column (3) for the latter and multiplying the quotient by 100. The series of columns detailing the percentage of other components to total working capital have been similārly obtained.

PROPORTION OF SHARE CAPITAL TO TOTAL WORKING CAPITAL

19. On reference to columns (2) and (3) of Table 10, it will be seen that agricultural societies in Bengal have proportionately much less share capital than in the rest of British India. The position is made clearer in column (4) which shows that Bengal had proportionately one-tenth of the share capital as compared with the rest of British India in the first year of the period under review viz., 1918-19. The position steadily improved till 1928-29 and 1929-30 when parity was reached, since when there has been a set back. The corresponding graphs will be found in Chart (1).

PROPORTION OF RESERVE FUND TO TOTAL WORKING CAPITAL

20. So far as reserve fund is concerned, Bengal's proportion is throughout higher than in the rest of British India as will be evident on comparing the figures in columns (5) and (6) in Table 10. It will, however, be improper to draw the facile conclusion from this that Bengal agricultural societies are stronger than those in the rest of British India. For, reserve funds show a steep rise after the onset of the Great Depression in Bengal as well as elsewhere and therefore they cannot indicate strength. Rather they presumably include items set aside for writing off bad and doubtful debts. This point will be further considered later on. The mutual position in Bengal and outside is illustrated in Chart (2).

PROPORTION OF MEMBERS' AND NON-MEMBERS' LOANS AND DEPOSITS TO TOTAL WORKING CAPITAL

21. Bengals' proportion is somewhat lower than in the rest of British India, but there is a substantial parallelism. The necessary figures are available in columns (8) to (10) for Members and in columns (11) to (13) for Non-members, the relative graphs being shown in Chart (3) and (4).

PROPORTION OF SOCIETIES' AND PROVINCIAL AND CENTRAL BANKS' LOANS AND DEPOSITS TO TOTAL WORKING CAPITAL

22. These loans and deposits form the bulk of the working capital and should therefore be closely scrutinised. It is somewhat curious that the higher financing bodies should curtail their accommodation in times of difficulty, as will appear from the steady decline in the proportion whether in Bengal or outside from the beginning of the Great Depression. There is since then substantial parallelism between the two as will appear from Chart (5).

PROPORTION OF OVERDUE LOANS TO TOTAL LOANS GRANTED BY AGRICULTURAL SOCIETIES

23. It has been stated above that probably the reserve funds of agricultural co-operative societies contain provision for bad and doubtful debts and their increase does not therefore necessarily indicate strength. An independent criterion has therefore to be devised for estimating the character of loans granted by such societies. The necessary data available in Table 11, columns (2), (3), (5) and (6) have been extracted from *Statistical Statements*, subject to a few adjustments in Bombay and U. P. figures[3]. Column (4) shows the percentage between columns (2) and (3) that is to say between overdue loans and total loans for Bengal. Column (7) gives the corresponding proportion for the rest of British India. If these two are mutually compared, it will be evident that the burden of overdue loans is a staggering one in Bengal rising to as much as 84·5% of the total in 1936-37. The position throughout has been worse than in the rest of British India and is almost desperate near the end of the period under review as will be evident from Chart (6). The corresponding figures for Bombay and Madras are shown in columns (8) and (9).

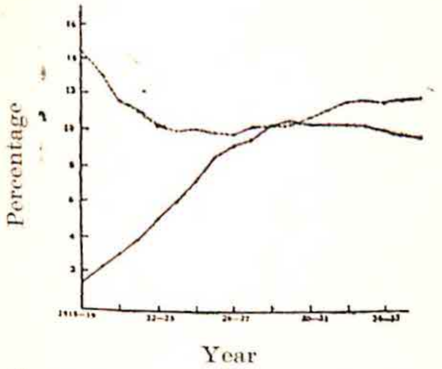


Chart (1). Percentage of Share Capital to Working Capital.

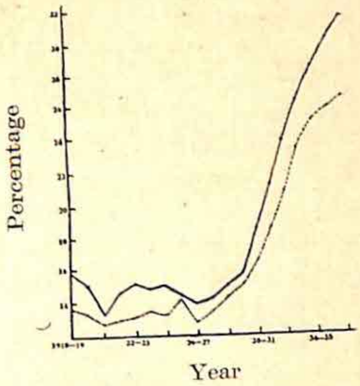


Chart (2). Percentage of Reserve Fund to Total Working Capital.

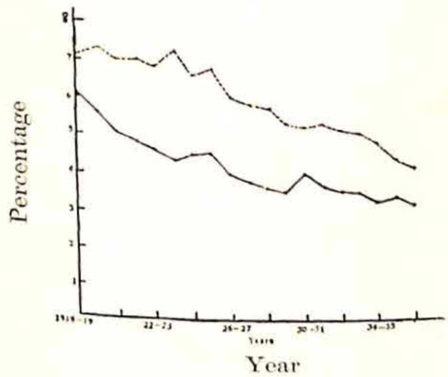


Chart (3). Percentage of Members' Loans and Deposits to Total Working Capital.

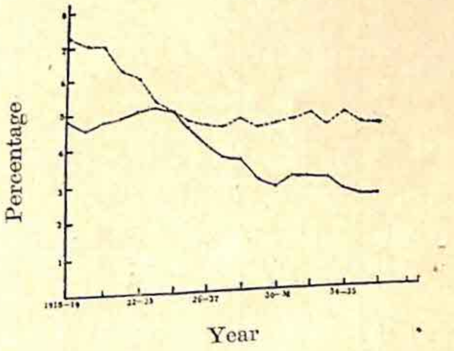


Chart (4). Percentage of Non-Members' Loans and Deposits to Total Working Capital.

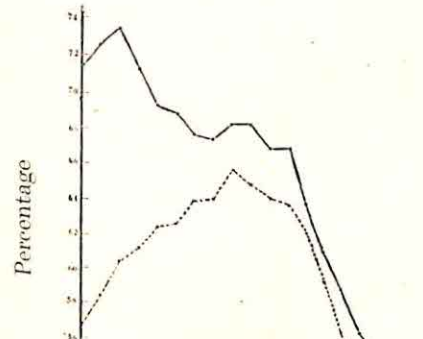


Chart (5). Percentage of Societies' Provincial and Central Banks' Loans and Deposits to Total Working Capital.

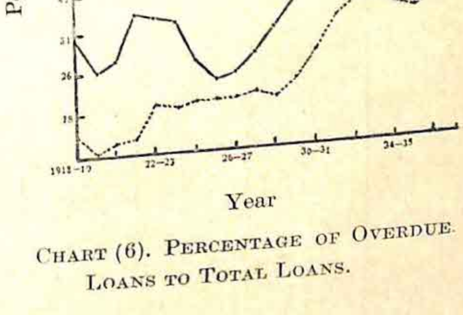


Chart (6). Percentage of Overdue Loans to Total Loans.

Bengal........ ——

Rest of India..................

TABLE 10. ANALYSIS OF WORKING CAPITAL OF AGRICULTURAL SOCIETIES.

year (July to June)	percentage of share capital to total working capital			percentage of reserve fund to total working capital			percentage of members' loans and deposits to total working capital			percentage of non-members' loans and deposits to total working capital			percentage of societies' provincial and central Banks' loans and deposits to total working capital			year (July to June)
	Bengal	rest of British India	disparity index 100(2)/(3)	Bengal	rest of British India	disparity index 100(5)/(6)	Bengal	rest of British India	disparity index 100(8)/(9)	Bengal	rest of British India	disparity index 100(11)/(12)	Bengal	rest of British India	disparity Index 100(14)/(15)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
1918—19	1·39	14·36	10	15·80	13·79	115	6·15	7·12	86	4·91	7·39	66	71·49	56·81	126	1918—19
—20	2·07	13·04	16	15·00	13·42	112	5·54	7·33	76	4·57	7·17	64	72·62	58·65	124	—20
—21	2·98	11·57	26	13·40	12·69	106	5·14	7·05	73	4·73	7·01	67	73·65	60·49	122	—21
—22	3·93	11·00	36	14·45	12·79	113	4·98	7·01	71	4·95	6·37	78	71·49	61·71	116	—22
—23	5·11	10·31	50	15·04	13·04	115	4·76	6·88	68	5·19	6·11	85	69·75	62·71	111	—23
—24	6·13	9·99	61	14·71	13·39	110	4·45	7·31	61	5·33	5·72	93	69·23	62·91	110	—24
—25	7·33	10·07	73	14·83	13·13	113	4·52	6·72	67	5·13	5·13	100	68·00	64·25	106	—25
—26	8·64	9·94	87	14·29	13·97	102	4·52	6·85	66	4·54	4·72	96	67·85	64·35	105	—26
—27	9·26	9·90	94	13·81	12·72	109	4·07	6·02	68	4·05	4·60	88	68·69	66·00	104	—27
—28	9·64	10·25	94	14·02	13·32	105	3·86	5·97	65	3·69	4·55	81	68·68	65·14	105	—28
—29	10·38	10·41	100	14·82	14·03	106	3·75	5·67	66	3·58	4·79	75	67·36	64·40	105	—29
—30	10·60	10·50	101	15·46	14·93	104	3·57	5·40	66	3·01	4·52	67	67·28	64·04	105	—30
—31	10·46	10·76	97	18·36	16·22	113	4·16	5·25	79	2·92	4·65	63	64·02	62·49	102	—31
—32	10·49	11·30	93	21·13	18·39	115	3·71	5·29	70	3·22	4·76	68	61·39	59·71	103	—32
—33	10·46	11·74	89	23·65	20·69	114	3·61	5·23	69	3·18	4·93	65	59·04	56·89	104	—33
—34	10·48	11·83	89	26·18	23·14	113	3·58	5·16	69	3·17	4·58	69	56·54	54·46	104	—34
—35	10·24	11·71	87	28·40	24·74	115	3·31	4·98	66	2·73	4·96	55	55·25	53·21	104	—35
—36	9·97	11·83	84	30·09	25·59	118	3·46	4·47	77	2·56	4·67	51	53·87	53·13	101	—36
—37	9·70	11·88	82	31·50	26·16	120	3·35	4·28	78	2·57	4·63	56	52·83	52·79	100	—37

TABLE 11. ANALYSIS OF LOANS-GRANTED TO INDIVIDUALS BY AGRICULTURAL SOCIETIES

Year (July to June)	Bengal			rest of British India			Bombay	Madras
	total loans to individuals Rs. lakhs	of which overdue Rs. lakhs	percentage of overdue loans to total loans 100(3)/(2)	total loans to individuals Rs. lakhs	of which overdue Rs. lakhs	percentage of overdue loans to total loans 100(6)/(5)	percentage of overdue loans to total loans	percentage of overdue loans to total loans.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1918—19	69 ·51	22 ·62	32 ·5	455 ·81	61 ·56	13 ·5	7 ·7	12 ·0
—20	85 ·34	22 ·08	25 ·9	554 ·97	59 ·66	10 ·8	8 ·4	11 ·8
—21	100 ·96	27 ·73	27 ·5	690 ·56	84 ·41	12 ·2	13 ·4	18 ·4
—22	106 ·81	40 ·05	37 ·5	792 ·87	99 ·84	12 ·6	13 ·5	19 ·2
—23	121 ·45	44 ·72	36 ·8	887 ·55	175 ·89	19 ·8	20 ·0	21 ·4
—24	148 ·47	51 ·78	34 ·9	995 ·76	191 ·70	19 ·3	19 ·7	24 ·7
—25	173 ·61	49 ·28	28 ·4	1223 ·09	243 ·08	19 ·9	17 ·9	24 ·2
—26	211 ·72	51 ·13	24 ·1	1398 ·39	281 ·51	20 ·1	16 ·3	30 ·9
—27	265 ·37	67 ·53	25 ·4	1540 ·46	302 ·77	19 ·7	23 ·7	28 ·1
—28	314 ·37	92 ·30	29 ·4	1837 ·63	386 ·75	21 ·0	33 ·4	26 ·8
—29	345 ·41	118 ·60	34 ·3	1994 ·78	403 ·14	20 ·2	31 ·1	26 ·8
—30	403 ·85	160 ·37	39 ·7	2144 ·16	494 ·81	23 ·1	30 ·7	30 ·1
—31	434 ·96	244 ·45	56 ·2	2198 ·85	638 ·24	29 ·0	29 ·6	44 ·5
—32	436 ·11	296 ·46	68 ·0	2068 ·66	749 ·08	36 ·2	44 ·7	55 ·4
—33	431 ·43	347 ·75	80 ·6	1958 ·12	768 ·44	39 ·2	49 ·3	59 ·1
—34	427 ·46	341 ·04	79 ·8	1885 ·23	745 ·55	39 ·5	46 ·7	60 ·3
—35	417 ·97	311 ·47	74 ·5	1847 ·39	709 ·95	38 ·4	48 ·3	53 ·9
—36	406 ·32	327 ·78	80 ·7	1850 ·89	676 ·28	36 ·5	54 ·5	46 ·8
—37	410 ·25	346 ·79	84 ·5	1828 ·24	637 ·82	34 ·9	49 ·4	39 ·5

TABLE 12. ANALYSIS OF LOANS AND DEPOSITS FROM PROVINCIAL AND CENTRAL BANKS TO AGRICULTURAL SOCIETIES

year (July to June)	Calcutta index numbers of wholesale prices	Bengal				rest of Br.-India				deflated value of loans and deposits per society	
		no. of agricultural societies (in 000's)	loans and deposits from Provin-cial and Central Banks (Rs. lakhs)			no. of agricultural Societies (in 000's)	loans and deposits from Provincial and Central Banks (Rs. lakhs).				
			actual value	deflated value	deflated value per society		actual value	deflated value	deflated value per society	Bombay	Madras
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1918—19	186	3 ·92	65 ·05	34 ·97	892	19 ·97	301 ·98	162 ·35	813	1,084	1,653
—20	205	4 ·98	80 ·11	39 ·07	785	28 ·81	378 ·71	184. 74	745	1,087	1,418
—21	188	5 ·85	93 ·54	49 ·76	851	28 ·87	480 ·73	255 ·71	886	1,351	1,448
—22	180	6 ·15	98 ·27	54 ·59	888	32 ·29	561 ·73	312 ·07	966	1,500	1,543
—23	173	7 ·21	108 ·25	62 ·57	868	34 ·51	640 ·29	370 ·11	1,072	1,602	1,663
—24	170	8 ·64	129 ·32	76 ·07	881	37 ·37	722 ·67	425 ·10	1,138	1,654	1,731
—25	168	10 ·20	150 ·68	89 ·69	879	41 ·51	873 ·02	519 ·06	1,250	2,081	1,744
—26	154	11 ·64	180 ·04	116 ·91	1,004	45 ·03	1027 ·10	666 ·95	1,481	2,661	1,956
—27	146	13 ·95	231 ·55	158 ·59	1,137	49 ·26	1246 ·68	853 ·89	1,733	3,037	2,276
—28	147	16 ·43	279 ·53	190 ·16	1,152	51 ·85	1360 ·20	925 ·31	1,785	2,845	2,414
—29	143	17 ·97	315 ·67	220 ·75	1,229	54 ·05	1456 ·50	1018 ·53	1,884	2,623	2,516
—30	132	20 ·44	370 ·83	280 ·93	1,375	55 ·42	1567 ·63	1187 ·60	2,143	3,180	2,887
—31	103	21 ·46	358 ·18	347 ·75	1,620	55 ·81	1574 ·90	1529 ·03	2,740	4,581	3,610
—32	94	21 ·49	354 ·01	276 ·60	1,752	55 ·56	1441 ·00	1532 ·98	2,759	4,949	3,438
—33	88	21 ·34	348 ·48	396 ·00	1,855	55 ·06	1328 ·12	1509 ·23	2,741	4,960	3,392
—34	89	21 ·21	334 ·36	375 ·69	1,771	54 ·92	1251 ·37	1406 ·04	2,560	4,665	3,180
—36	89	21 ·12	334 ·21	375 ·51	1,778	55 ·67	1226 ·32	1377 ·88	2,475	4,490	3,314
—36	91	21 ·11	328 ·97	361 ·51	1,712	56 ·98	1249 ·69	1373 ·28	2,410	4,167	3,644
—37	96	21 ·26	324 ·24	337 ·75	1,589	58 ·75	1249 ·40	1301 ·46	2,215	4,031	3,739

PROPORTION OF LOANS AND DEPOSITS FROM PROVINCIAL AND CENTRAL BANKS PER AGRICULTURAL SOCIETY

24. Reference has been made above to the stringent loan policy of higher financing institutions in times of crisis in the case of agricultural societies. Co-operative central banks are not "lenders in the last resort" like the commercial central banks of the money market : they are also lenders in the first resort. The former cannot therefore follow a liberal loan policy like the latter in times of difficulty. But the question is, have the higher financing institutions been more exacting in Bengal than elsewhere, leading to the deplorable position described above ? For answering this question, reference has to be made to Table 12 columns (3), (4), (7) and (8). The other columns are self-explanatory. On comparing column (6) for Bengal with column (10) for the rest of British India, it will be seen that the latter have followed a more liberal loan policy than the former. The niggardly treatment in Bengal as compared with Bombay and Madras is apparent from columns (6), (11) and (12).

SUMMARY AND CONCLUSIONS

It will be convenient to sum up the main conclusions below :—

(1) Data as published in the official *Statements Relating to the Co-operative Movement in India* require some adjustment before being used (paragraphs 6-9), the adjusted data being shown in Tables 2 and 3.

(2) There is some difference in the rates of progress of agricultural and non-agricultural societies, whether in Bengal or in the rest of British India (paragraph 11).

(3) It is only when the working capital is deflated by an appropriate index number of prices that this difference is revealed (paragraph 12).

(4) Progress of co-operation in relation to population has been such that trends need not be computed by Gompertz or Logistic Curves (paragraph 13).

(5) If logarithmic linear trends are fitted and the *t*-test applied, Bengal's rate of growth is higher than that of the rest of British India in the case of (a) number of societies, both agricultural and non-agricultural, (b) membership of both of these two types of societies, and (c) working capital, both ordinary and deflated, of agricultural societies, whereas in the case of working capital of non-agricultural societies, whether ordinary or deflated, the rate of growth in Bengal is the same as in the rest of British India (paragraph 14, Table 6).

(6) Agricultural societies have fewer members in Bengal than in the rest of British India. In the case of membership of non-agricultural societies also, Bengal has failed to keep pace with Bombay and Madras (paragraph 15).

(7) So far as working capital per society is concerned, Bengal's position is hopelessly inferior in the case of agricultural societies, and somewhat so in the case of non-agricultural societies (paragraph 16).

(8) Even with respect to working capital per member, the position in Bengal is definitely inferior to that in Bombay and Madras, whether for agricultural or for non-agricultural societies (paragraph 17).

(9) To secure comparable standards, the several components of the total working capital of agricultural societies have been expressed as percentages of the total (paragraph 18) leading to the following conclusions :—

(a) Bengal has proportionately less share capital (paragraph 19) ;

(b) but higher reserve fund, probably because reserves for bad and doubtful debts are also included in the reserve fund (paragraph 20) ;

(c) the proportion of loans and deposits from members and non-members is also somewhat lower in Bengal (paragraph 21) ;

(d) the proportion of loans and deposits from societies, Provincial and Central Banks is also different (paragraph 22).

(10) The real nature of the reserve fund in Bengal is revealed by the staggering percentage of overdue loans to total loans, rising from 24·1% in 1925-26 to as much as 84·5% in 1936-37 (paragraph 23).

(11) The higher financing institutions in Bengal have followed a more niggardly policy in the matter of helping the primary societies out of the crisis than in the rest of British India (paragraph 24).

REFERENCES

1. For a copy of the earliest balance sheet and report of this earliest co-operative society in Bengal see "Co-operation in Bengal" by H. Sinha in the *Visva-Bharati Quarterly*, Vol. 8, parts 1 and 2, pp.99-100.

2. All annual reports belonging to this series have been referred to as *Statistical Statements* in the paper, whether before 1927-28 or not.

3. In the *Statistical Statements* for 1926-27, p.13, General Statement B, the total loan due by individuals to agricultural societies in Bombay during that year is shown as Rs. 2,16,70,325 whereas the amount of overdue loan (which must obviously be less) is shown as Rs. 5,12,95,110. As reference to the Registrar of Co-operative Societies, Bombay, did not bring any reply, the latter figure has been tentatively corrected to Rs. 51,29,511. The Registrar of Co-operative Societies, United Provinces, however, kindly corrected the figure for total loans due by individuals during 1931-32 as published in the *Statistical Statements*. As directed by him, Rs. 30,00,000 has been added to the published figure. See Prof. J. C. Sinha's *Indian Currency Problems during the past Decade*, p.100n.

4. The All-India index number of wholesale prices published in the official publication *Index Numbers of Prices in India* (Base 1873=100) does not at all represent the present trend of wholesale prices in India. Moreover, only calendar year figures are available for the period prior to the publication of *Monthly Survey of Business Conditions in India* which started the monthly series. It is therefore impossible to calculate index numbers for co-operative years ending June 30 for earlier years.

The fact is that there are several wholesale markets in India, where different wholesale prices are current, and are of varying importance, for which separate index numbers have to be computed. They cannot, however, be combined into an all-India one. We agree with Prof. J. C. Sinha in his view that the Calcutta index is the most representative for the reasons stated by him in his *Indian Currency Problems During the past Decade*, pp. 16-17. For a further discussion on this point see "The Ratio Question" by Drs. H. Sinha and J. C. Sinha in *Sankhya : The Indian Journal of Statistics*. Vol. 3, part 2, pp.126-127, Mr. B. P. Adarkar's "Ratio Question—A Criticism" in *Sankhya : The Indian Journal of Statistics*, Vol. 3, part 4, .378 and Drs. H. Sinha and J. C. Sinha's "The Ratio Question—A Reply" in *Sankhya : The Indian Journal of Statistics*, Vol. 3, part 4, pp. 394-395.

5. It should , however, be remembered that although the deflated values are shown up to the units place in this and in the following tables, the figures are generally correct only up to the hundreds place, on account of divisions and multiplications. They are, however, written in full to facilitate comparison with the raw unadjusted data.

6. "Law of Growth in forecasting Demand" by R. B. Prescott in the *Journal of the American Statistical Association*, December, 1922, pp.471-479, Vol. XVIII.

7. The formula for comparing the slopes of the two regression equations is given in Mills: *Statistical Methods* (Revised Edition—1938) pp. 608-610.

If $y_1 = a_1 + b_1 t_1$ and $y_2 = a_2 + b_2 t_2$ be the two regression equations then $t = (b_1 - b_2)/\sigma b_1 - b_2$

Where $\sigma_{b_1 - b_2}$ = the standard error of the difference between b_1 and b_2

$$= \sqrt{S^2_y/\Sigma x^2_1 + S^2_y/\Sigma x^2_2}$$

and S^2_y = measure of the average scatter about the two lines of regression

$$= \{\Sigma(y_1 - yc_1)^2 + \Sigma(y_2 - yc_2)^2\}/(N_1 + N_2 - 4)$$

yc_1 being the calculated value of y_1

yc_2 " " " y_2

N_1 " number of degrees of freedom of the first equation

N_2 " number of degrees of freedom of the second equation.

Paper received : 1 December, 1941.

ON THE TOTALITY OF THE SOLUTIONS FOR THE SYMMETRIAL INCOMPLETE BLOCK DESIGNS : $\lambda=2$, k= 5 or 6.

By Q. M. HUSAIN
Dacca University

INTRODUCTION

The combinatorial problem involved in the construction of balanced incomplete block designs has been solved in most of the useful cases and in various ways. Fisher and Yates (1938) have embodied in their table a number of such solutions ; Bose (1939) has tried to develop a systematic procedure for constructing such designs while Carmichael (1937) gives various ingenious methods for particular cases. But the problem of testing the isomorphism of the various solutions suggested for any single design remains to be attacked, as also the problem of investigation of the totality of non-isomorphic solutions. In the present paper all possible non-isomorphic solutions of the two designs, (i) $v=b=11$, $r=k=5$, $\lambda=2$ and (ii) $v=b=16$, $r=k=6$, $\lambda=2$ have been enumerated by following a standard method of building up a design step by step which may be briefly explained in the case of (ii) as follows : The first block is filled up with the six standard treatments 1, 2, ..., 6. Since in the remaining fifteen blocks each of the $\binom{6}{2}$ or fifteen pairs must occur, we place these pairs in the fifteen blocks. The new varieties 7, 8, ...,16 are now to be placed. Each of them will occur with either of the series of blocks (ab) (bc) (cd) (de) (ef) (fa) or (ab), (bc), (ac), (de) (ef) (df) where a, b, c, d, e, f are the six initial treatments which may be called in shortened form the λ-chain (abcdef) or the λ-chain (abc)(def) respectively. Taking all possible λ-chains into consideration it has been possible to enumerate the non-isomorphic solutions for these symmetrical designs. It has been found that,

(i) the design $v=b=11$, $r=k=5$, $\lambda=2$ has one and only one solution with which all other solutions suggested for this case are isomorphic.

(ii) the design $v=b=16$, $r=k=6$, $\lambda=2$ has three non-isomorphic solutions which are given in appendix I. The designs given by Fisher and Yates (1938) or Bose (1939) or the first solution of Carmichael (1937) are also isomorphic with the first solution given there, while the second solution of Camichael (1937) is isomorphic with the third solution ; the method of identification and test of isomorphism being explained in appendix II.

2. THE DESIGN $v=b=11$, $r=k=5$, $\lambda=2$

The initial block contains the varieties 1,2, 3, 4, 5. Each of the remaining 10 blocks must contain one of the 10 pairs of the varieties 1, 2,..., 5. So we have now to place only the six varieties 6, 7..., 11 to complete the design. A necessary condition is that each of the varieties 6,...,11 must occur twice with the varieties in the initial block. This can be readily represented by the λ-chaining of the type (12345), meaning thereby that one of these varieties, say 6, must occur in the blocks containing (1,2) ; (2,3) ; (3,4) ; (4.5) and (5,1), each number being repeated λ times. This is the only type of λ-chaining or cycle possible with only five elements.

Supposing '6' to be placed in the chain (12345) we can place two more varieties say 7 and 8, in the block (1,2), the other possible chains containing the varieties 1 and 2, are (12354), (12435), (12453), (12534) and (12543) of which the first, second and the fifth are inconsistent with (12345). This is obvious from the the fact that any two chains must have just two pairs of varietal numbers common. If more than two pairs of numbers are common, this means that the two varieties are common to more than two blocks. Thus, three varieties, say 6, 7, 8 are placed in any order in the chains :

(12345) ; (12453) ; (12534).

Now at this stage, the block (1,3) has already received three varieties. Two more varieties can be placed in it. The chaining of these two varieties say 9 and 10, must be consistent with (12453) or (13542) that has been already used. The possible chains other than this are :

(13524) ; (13425) ; (13452) ; (13254) : and (13245)

of which the first, the third, and the fifth are inconsistent with (13542).

Hence, the only possible chains for placing these two varieties (in any order) are ; (13425), (13254) which are also consistent with the chains already used. The remaining variety 11 is then uniquely placed in the chain (14235) as is easily seen. The complete pattern may be represented by the standard placing of the varieties 6 to 11 in the unique set of chains ;

(12345), (12453), (12534), (13425), (13254),and (14235)

Regarding the successive chains to belong to 6, 7, 8, 9, 10, 11 respectively, the completed pattern is shown below. Of course, the actual order in which these varieties are allotted to the successive chains is immaterial.

(1, 2, 3, 4, 5)
(1, 2, 6, 7, 8)
(1, 3, 7, 9, 10)
(1, 4, 8, 10, 11)
(1, 5, 6, 9, 11)
(2, 3, 6, 10, 11)

(2, 4, 7, 9, 11)
(2, 5, 8, 9, 10)
(3, 4, 6, 8, 9)
(3, 5, 7, 8, 11)
(4, 5, 6, 7, 10)

3. THE DESIGN $v=b=16$, $r=k=6$, $\lambda=2$

Naming the varieties in the initial block as 1, 2, 3, 4, 5, 6 and denoting the other blocks by the 15 pairs of these initial numbers, we can represent the blocks in which the varieties 7 to 16 are placed, by two types of λ-chaining, viz, A=(123456) and B=(123)(456) which are shortened ways of writing the blocks (1,2) (2,3) (3,4) (4,5) (5,6), (6,1) in type A and the blocks (1,2) (2,3) (3,1) (4,5) (5,6) (6,4) in type B. No other type of λ-chaining is possible with six varieties.

Type B comprises the following 10 chains :—

(1 2 3) (4 5 6)
(1 2 4) (3 5 6)
(1 2 5) (3 4 6)
(1 2 6) (3 4 5)
(1 3 4) (2 5 6)

(1 3 5) (2 4 6)
(1 3 6) (2 4 6)
(1 4 5) (2 3 6)
(1 4 6) (2 3 5)
(1 5 6) (2 3 4)

We notice that the set of numbers within brackets may be regarded as an unreduced solution of the design $v=6$, $b=20$, $r=10$, $k=3$, $\lambda=4$ and that the whole set breaks up into 10 pairs each representing a complete replication, and a λ-chain of type B. We readily see that in this set of the λ-chains,

(i) Each variety (from 7 to 16) occurs in six blocks, satisfying $r = 6$.

(ii) Each pair of initial varieties occurs four times, i.e., receives four of the varieties from 7-16, thus satisfying $k=6$.

(iii) Between any two λ-chains there are two pairs of blocks common showing that every variety pair occurs exactly in two blocks, i.e. the condition $\lambda=2$ is satisfied. Also for the whole design, $v=b=16$ is satisfied.

Thus the set of λ-chains of type B provide a solution of the design under examination. It can be easily seen that the design given by Fisher & Yates (1938) and R. C. Bose (1939) are isomorphic with this solution.

We next examine the possibility of a design in which one or more varieties are placed in λ-chains of type A. Without loss of generality we can write one of the chains as (123456). The totality of chains consistent with this is as follows :—

Type A.

(1 2 4 6 5 3) ; (1 2 5 4 6 3)
(1 2 6 3 5 4) ; (1 2 6 4 3 5)
(1 2 6 4 5 3) ; (1 3 2 4 6 5)
(1 3 2 5 6 4) ; (1 3 2 6 4 5)
(1 3 4 2 6 5) ; (1 3 5 4 2 6)
(1 4 2 3 6 5) ; (1 4 3 5 2 6)
(1 5 2 4 3 6) ; (1 5 3 2 4 6)
(1 5 3 4 2 6)

Type B.

(1 2 4) (3 5 6)
(1 2 5) (3 4 6)
(1 3 4) (2 5 6)
(1 3 5) (2 4 6)
(1 3 6) (2 4 5)
(1 4 5) (2 3 6)
(1 4 6) (2 3 5)

We notice that one chain of type A precludes three of type B.

The four varieties necessary to complete the block (1,2) can have if possible the following combinations of λ-chains.

	(a)	(b)	(c)
Type A.	4	3	2
Type B.	0	1	2
Total	4	4	4

Let us try (a) with 4 chains of type A.

On examining the five chains of type A containing the variety pair (1,2) and each consistent with the chain (123456) we find that no three of them are consistent among themselves. There are only two consistent pairs, viz.,

(124653) and (125463)
(126354) (126435)

This shows that (a) is not possible ; but (b) and (c) are possible, provided the chains of type B can be found consistent with those used from type A.

The cases arising, so far as the four varieties in the block (1,2) are concerned, are shown below with appropriate placings of chains of both the types :

(α)	(β)	(γ)
(1 2 3 4 5 6)	(1 2 3 4 5 6)	(1 2 3 4 5 6)
(1 2 4 6 5 3)	(1 2 5 4 6 3)	(1 2 6 4 5 3)
(1 2 6 3 5 4)	(1 2 6 4 3 5)	(1 2 4) (3 5 6)
(1 2 5) (3 4 6)	(1 2 4) (3 5 6)	(1 2 5) (3 4 6)

We notice that in each of the sets (α), (β) and (γ) one variety has been placed in the block (1,3), so that three more varieties are to be placed in this block.

In the set (α) only one of the chains of type A containing the variety pair (1,3) is consistent with the chain already used in the block (1,3). This is consistent also with the other placings of (α). We need therefore. two chains of type B to complete the block (1,3). The set of chains up to this are :

(α)	(1 2 3 4 5 6)	(1 3 2 6 4 5)	(1 4 2 3 6 5)	(1 5 3 4 2 6)
	(1 2 4 6 5 3)	(1 3 4) (2 5 6)	(1 4 6) (2 3 5)	
	(1 2 6 3 5 4)	(1 3 6) (2 4 5)		
	(1 2 5) (3 4 6)			

The chains of type B are best selected by considering those precluded by the chains of type A used in the set. This is the only way, as all the three chains cannot be of type B. At this stage the block (1,4) has received two varieties. So it is necessary to place two more varieties in this block. It is seen that both the chains cannot be of type B, nor can both be of the type A. The only consistent chain of type A is (142365) and the chain of type B is (146) (235). The combinations are shown under (α).

At this stage only one variety remains to be placed. There are no chains of type B available. The only consistent chain of type A is shown under (α) giving a complete solution. This contains 4 chains of type B. It is note-worthy that the set (α) can be completed only in one way as shown above.

We next consider the set (β). Three varieties are to the placed in the block (1,3). The chain (125463) has been already used. This precludes the chain (136) (245) so all the three chains cannot be of type B. The two chains of type A consistent with the earlier placings are (132465) and (135426) which are consistent between themselves. Of the two available chains of type B, (135) (246) is not consistent with either of the chains of type A. Thus, the only possibility is the following shown under (β).

(β)	(1 2 3 4 5 6)	(1 3 2 4 6 5)	(1 4 5) (2 3 6)	(1 5 2 4 3 6)
	(1 2 5 4 6 3)	(1 3 5 4 2 6)	(1 4 6) (2 3 5)	
	(1 2 6 4 3 5)	(1 3 4) (2 5 6)		
	(1 2 4) (3 5 6)			

At this stage, the block (1,4) needs two more varieties to be completed. We are to choose two out of the four chains :

(1 4 5) (2 3 6)	(1 4 2 3 6 5)
(1 4 6) (2 3 5)	(1 4 3 5 2 6)

The two chains of type A are not consistent with each other. So we have to try either two cycles of type B or one each of types A and B. Neither of the chains of Type A is consistent with all the chains previously used. Hence, we are left with only the two cycles of type B. The last variety is then placed uniquely in the chain (152436) thus completing the solution of the set (β). This contains 4 chains of type B.

Finally let us consider the set (γ). Three varieties are to be placed in the block (1,3). The chain (126453) has been already used. This precludes the set (135) (246) of the type B. Hence, all the three chains cannot be of type B. The following chains are consistent with the earlier placings :

(1 3 2 4 6 5)	(1 3 4) (2 5 6)
(1 3 2 5 6 4)	(1 3 6) (2 4 5).
(1 3 4 2 6 5)	

The choice of three chains out of these can be made in two ways only, viz, either, one of type A (with two of type B) or two of type A (with one of type B). The possible combinations, upto this stage are shown below under (γ_1) and (γ_2);

(γ_1)	(1 2 3 4 5 6) (1 2 6 4 5 3) (1 2 4) (3 5 6) (1 2 5) (3 4 6)	(1 3 2 4 6 5) (1 3 4) (2 5 6) (1 3 6) (2 4 5)	(1 4 5) (2 3 6) (1 4 6) (2 3 5) (1 5 3 4 2 6)
(γ_2)	(1 2 3 4 5 6) (1 2 6 4 5 3) (1 2 4) (3 5 6) (1 2 5) (3 4 6)	(1 3 2 5 6 4) (1 3 4 2 6 5) (1 3 6) (2 4 5)	(1 4 3 5 2 6) (1 4 5) (2 3 6) (1 5 3 2 4 6)

In (γ_1) the varieties in block (1,4) form into two chains of type B. In (γ_2) they form into one of type A and one of type B. The block (1,5) receives the remaining variety which is uniquely placed in either of the designs: (γ_1) and (γ_2) contain respectively, 6 and 4 chains of type B.

One important fact is that the number of chains of either type in a solution, can never be altered by any renaming of the varieties. Hence, solutions containing an unequal number of chains of type A (or of type B) cannot be isomorphic. We have then obtained at least three non-isomorphic solutions viz. those containing 10,6 and 4 chains of type B. Of the 5 solutions shown, (α), (β) and (γ_2) contain 4 chains of type B. These chains are shown below:

(α)	(β)	(γ_2)
(1 2 5) (3 4 6)	(1 2 4) (3 5 6)	(1 2 4) (3 5 6)
(1 3 4) (2 5 6)	(1 3 4) (2 5 6)	(1 2 5) (3 4 6)
(1 3 6) (2 4 5)	(1 4 5) (2 3 6)	(1 3 6) (2 4 5)
(1 4 6) (2 3 5)	(1 4 6) (2 3 5)	(1 4 5) (2 3 6)

It is necessary to test whether the corresponding solutions are isomorphic. It is seen that (β) is changed to (α) by the transpositions (15) (42) and (γ_2) is changed to (α) by the transpositions (26) (35). These transpositions make not only the double chains but the complete set of chains identical as may be readily verified. When the set of chains are identical the complete identity of patterns is only a matter of renaming the non-initial varieties.

Thus we have finally proved that the design: $v=b=16$, $r=k=6$, and $\lambda=2$ admits of 3 distinct types of solutions, leading to 3 non isomorphic designs. The three complete designs are shown in appendix I.

APPENDIX I (THE DESIGNS)

The complete patterns corresponding to the three fundamental sets of λ-chains are recorded below for reference. The sets I, II and III represent respectively, 10,5 and 4 chains of type B. For II and III, the solutions (γ_1) and (γ_2) are recorded.

I	II	III
(1, 2, 3, 4, 5, 6)	(1, 2, 3, 4, 5, 6)	(1, 2, 3, 4, 5, 6)
(1, 2, 7, 8, 9, 10)	(1, 2, 7, 8, 9, 10)	(1, 2, 7, 8, 9, 10)
(1, 3, 7, 11, 12, 13)	(1, 3, 8, 11, 12, 13)	(1, 3, 8, 11, 12, 13)
(1, 4, 8, 11, 14, 15)	(1, 4, 9, 12, 14, 15)	(1, 4, 9, 11, 14, 15)
(1, 5, 9, 12, 14, 16)	(1, 5, 10, 11, 14, 16)	(1, 5, 10, 12, 15, 16)
(1, 6, 10, 13, 15, 16)	(1, 6, 7, 13, 15, 16)	(1, 6, 7, 13, 14, 16)
(2, 3, 7, 14, 15, 16)	(2, 3, 7, 11, 14, 15)	(2, 3, 7, 11, 15, 16)
(2, 4, 8, 12, 13, 16)	(2, 4, 9, 11, 13, 16)	(2, 4, 9, 12, 13, 16)
(2, 5, 9, 11, 13, 15)	(2, 5, 10, 12, 13, 15)	(2, 5, 10, 11, 13, 14)
(2, 6, 10, 11, 12, 14)	(2, 6, 8, 12, 14, 16)	(2, 6, 8, 12, 14, 15)
(3, 4, 9, 10, 11, 16)	(3, 4, 7, 10, 12, 16)	(3, 4, 7, 10, 12, 14)
(3, 5, 8, 10, 12, 15)	(3, 5, 8, 9, 15, 16)	(3, 5, 8, 9, 14, 16)
(3, 6, 8, 9, 13, 14)	(3, 6, 9, 10, 13, 14)	(3, 6, 9, 10, 13, 15)
(4, 5, 7, 10, 13, 14)	(4, 5, 7, 8, 13, 14)	(4, 5, 7, 8, 13, 15)
(4, 6, 7, 9, 12, 15)	(4, 6, 8, 10, 11, 15)	(4, 6, 8, 10, 11, 16)
(5, 6, 7, 8, 11, 16)	(5, 6, 7, 9, 11, 12)	(5, 6, 7, 9, 11, 12)

It may be noted that I and II have 6 chains in common (all of type B)
I and III have 4 chains in common (all of type B)
II and III have 6 chains in common (4 of type B and 2 of type A)

This shows that the identity of even 6 chains of type B does not necessarily lead to the same pattern.

If p consistent chains of type A can be found which exclude exactly 'p' chains of type B, then the 'p' chains of type A and the (10-p) chains of type B together form a consistent set corresponding to a solution of the design. The above result shows that 'p' can have only the values 0,4 and 6.

APPENDIX II (THE PROBLEM OF IDENTIFICATION)

It is of importance to be able to check whether two given designs with the same parameters v, b, r, k, λ are isomorphic. As an example let us see how we can quickly arrive at a substitution that would make Fisher's and Bose's solutions identical if possible.

If we regard the quadruplet treatment in Bose's design as a number expressed in the scale of 2 and call (0000) as 16; further if we name the sixteen alphabets in Fisher's design from 1 to 16, the two solutions can be respectively written as:

(B)	(F)
(16. 1. 8, 15, 5, 3)	(1, 2, 3, 4, 5, 6)
(1, 16, 9, 14, 4, 2)	(1, 2, 7, 8, 9, 10)
(2, 3, 10, 13, 7, 1)	(1, 3, 7, 11, 12, 13)
(3, 2, 11, 12, 6, 16)	(1, 4, 8, 11, 14, 15)
(4, 5, 12, 11, 1, 7)	(1, 5, 9, 12, 14, 16)
(5, 4, 13, 10, 16, 6)	(1, 6, 10. 13, 15, 16)
(6, 7, 14, 9, 3, 5)	(2, 3, 7, 14, 15, 16)
(7, 6, 15, 8, 2, 4)	(2, 4, 8, 12, 13, 16)
(8, 9, 16, 7, 13, 11)	(2, 5, 9, 11, 13, 15)
(9. 8, 1, 6, 12, 10)	(2, 6, 10, 11, 12, 14)
(10, 11, 2, 5, 15, 9)	(3, 4, 9, 10, 11, 16)
(11, 10, 3. 4, 14, 8)	(3, 5, 8, 10, 12, 15)
(12, 13, 4, 3, 9, 15)	(3, 6, 8. 9, 13, 14)
(13, 12, 5, 2, 8, 14)	(4, 5, 7, 10, 13, 14)
(14, 15, 6, 1, 11, 13)	(4, 6, 7, 9, 12, 15)
(15, 14, 7, 16, 10, 12)	(5, 6, 7, 8, 11, 16)

In the present case F is in standard form, but this is not a necessary requirement. The following steps may be followed :—

(1) Identify any block of F with any of B. We have chosen the two first blocks. Identify the varieties occurring in these blocks (say, initial blocks) in any order. We have preserved the order in which they have been written. This gives the partial substitution : $\begin{pmatrix} 1, 2, 3, 4, 5, 6 \\ 16, 1, 8, 15, 5, 3 \end{pmatrix}$

(2) Identify the blocks containing equivalent initial pairs, i.e. blocks (1,2), (1,3)... of F with blocks (16,1) (16,8)...of B.

(3) Form the λ-chains of the non-initial varieties of both the given designs. If these are identical, then the varieties corresponding to identical chains are transposable. But it may be that the chains are not identical. In this case make one of the chains for one design conform to some chain of the other design by a substitution. If the two solutions are isomorphous the two complete sets of chains will become identical by the same substitution. Having got the identical chains, set up the equivalence of the varieties that correspond to the same chain.

Thus, in general we have two sets of substitutions one for the initial varieties and the other for the non-initial. These substitutions bring the two designs to identity, if this is at all possible.

REFERENCES

1. Bose, R. C. (1939) On the construction of balanced incomplete block designs. *Annals of Eugenics*, **9**, 353-399.

2. Carmichael, R. D. (1937) Introduction to Groups of Finite Order, Ginn & Co. London.

3. Fisher, R. A. & Yates, F. (1938) Statistical Tables for Biological, Agricultural and Medical Research, Oliver & Boyd, London & Edinburgh.

Paper received : 16 *September*, 1943.

MISCELLANEOUS NOTES

BERNOULLI'S THEOREM AND TSHEBYCHEFF'S ANALOGUE

By SAMARENDRA NATH ROY & PURNENDU BOSE
Statistical Laboratory

Bernoulli's theorem on the probability that in a series of n trials (where the chance of success is constant from trial to trial and equals p) the number of successes m should be within a certain range of the expectation value, can be stated as

$$P\{|m/n-p|<\varepsilon\}>1-\eta \quad \text{if } n>n_0(\varepsilon,\eta) \quad .. \quad (1\cdot1)$$

where ε and η are arbitrary small numbers and

$$n_0(\varepsilon,\eta)=(1/\varepsilon)+(1+\varepsilon)\,log(1/\eta)/\varepsilon^2. \quad .. \quad (1\cdot2)$$

By introducing n independent stochastic variables $x_1, x_2, \ldots, x_n$ each of which can take values 1 or 0 with probabilities p or q we can deduce from Tshebycheff's law of large numbers a theorem which is of the same form as Bernoulli's theorem (1·1), except that $n_0(\varepsilon, \eta)$ is a function of ε and η different from that given by (1.2). This theorem we shall call Tshebycheff's analogue of Bernoulli's theorem. In its general form Tshebycheff's law of large numbers can be stated as

$$P\{|(x_1+x_2+\ldots+x_n)/n-E(x_1+x_2+\ldots+x_n)/n|<\varepsilon\}>1-\frac{V(x_1+x_2+\ldots+x_n)/\varepsilon^2}{n} \quad .. \quad (2\cdot1)$$

where $x_1, x_2, \ldots, x_n$ are any general stochastic variables, $E(x_1+x_2+\ldots+x_n)$ the mathematical expectation and $V(x_1+x_2+\ldots+x_n)$ the variance of $x_1+x_2+\ldots+x_n$, both being supposed to exist. If now, instead of $x_1, x_2, \ldots, x_n$ being general stochastic variables, they are assumed to be mutually independent stochastic variables such that each can take on values 1 or 0 with probabilities p or q then

$$\left.\begin{aligned}(x_1+x_2+\ldots+x_n) &= m(\text{number of successes})\\ E(x_1+x_2+\ldots+x_n)/n &= p\\ V(x_1+x_2+\ldots+x_n)/n &= pq/n\end{aligned}\right\} \quad .. \quad (2\cdot2)$$

It will be thus seen that by utilising (2.2), (2.1) can be reduced to

$$P\{|m/n-p|<\varepsilon\}>1-pq/n\varepsilon^2 \quad .. \quad (2\cdot3)$$

Since p and q are both positive, and $p+q=1$, the maximum value of pq is reached when $p=q$ and that value is $\frac{1}{4}$. Hence (2.3) implies another inequality relation

$$P\{|m/n-p|<\varepsilon\}>1-1/4n\varepsilon^2 \quad .. \quad (2\cdot4)$$

It follows that given two arbitrary small numbers ε and η we have

$$P\{|m/n-p|<\varepsilon\}>1-\eta \quad \text{if} \quad n\geqslant n'_0(\varepsilon,\eta) \quad .. \quad (2\cdot5)$$

where $n'_0(\varepsilon, \eta) = 1/4\varepsilon^2\eta$. Relation (2.5) we shall call Tshebycheff's analogue of Bernoulli's theorem, which latter would be defined by (1.1) and (1.2). Now both Bernoulli's theorem and Tschebycheff's analogue are approximate and fairly close but not exact inequality relations, that is, in neither case would the inequality relation (1.1) or (2.5) be necessarily reversed if $n\leqslant n_0(\varepsilon, \eta)$ defined by (1.2) or $n\leqslant n'_0(\varepsilon, \eta)$ defined by (2.5). There is, of course, a least number $n''_0(\varepsilon, \eta)$ which is yet unknown such that the inequality relation (1.1) or (2.5) holds if $n>n''_0(\varepsilon, \eta)$ and is reversed if $n<n''_0(\varepsilon,\eta)$. The question now naturally arises as to which of $n_0(\varepsilon, \eta)$ or $n'_0(\varepsilon, \eta)$ is nearer to $n''_0(\varepsilon, \eta)$ for different values of ε and η; both are, of course, greater than $n''_0(\varepsilon, \eta)$. This point we have tried to settle by numerical-cum-graphical methods in the following manner.

Considering the (ε, η) plane we are generally interested in the square defined by ε varying from 0 to 1, and η varying from 0 to 1. Inside this square we are interested in knowing which of $n_0(\varepsilon, \eta)$ or $n'_0(\varepsilon, \eta)$ is greater. We have thus to find whether $1/\varepsilon+(1+\varepsilon)\,log(1/\eta)/\varepsilon^2$ or $1/4\varepsilon^2\eta$ is greater.

If $n_o\ (\varepsilon, \eta)$ is greater, then Bernoulli's theorem gives a wider limit and is thus worse, and if $n'_o\ (\varepsilon, \eta)$ is greater it is the other way about. Equating the two we have

$$1/\varepsilon+(1+\varepsilon)\ log\ (1/\eta)/\varepsilon^2=1/4\varepsilon^2\eta \qquad .. \quad (3\cdot1)$$

or

$$\varepsilon=\{1/4\eta-log\ (1/\eta)\}/\{1+log\ (1/\eta)\} \qquad .. \quad (3\cdot2)$$

If this curve had cut the (ε, η) plane outside the square defined above, then within our region of interest either $n_o\ (\varepsilon, \eta)$ or $n'_o(\varepsilon, \eta)$ would have been throughout the greater of the two, that is, either Tshebycheff's analogue or Bernoulli's theorem would have been better. But, as the accompanying table and graph would show, this curve (3.2) cuts across the square and divides it into two portions in one of which $n_o(\varepsilon,\eta)>n'_o(\varepsilon, \eta)$ and in the other $n'_o(\varepsilon, \eta)>n_o\ (\varepsilon.\eta)$; that is, in one region Tshebycheff's analogue is closer and in the other Bernoulli's theorem is closer.

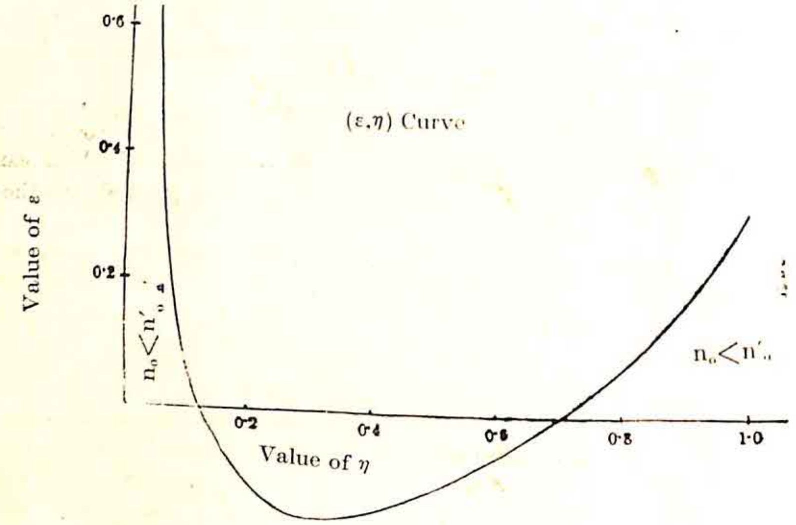


TABLE 1. VALUES OF ε FOR DIFFERENT VALUES OF η

η	ε	η	ε	η	ε	η	ε
0·01	3·6386	0·09	0·1085	0·17	−0·1087	0·50	−0·1141
0·02	1·7484	0·10	0·0598	0·18	−0·1200	0·55	−0·0897
0·03	1·0710	0·11	0·0204	0·19	−0·1296	0·60	−0·0623
0·04	0·7185	0·12	−0·0118	0·20	−0·1377	0·70	0·0003
0·05	0·5016	0·13	−0·0385	0·25	−0·1619	0·80	0·0731
0·06	0·3549	0·14	−0·0608	0·30	−0·1682	0·90	0·1560
0·07	0·2493	0·15	−0·0795	0·40	−0·1520	1·00	0·2500
0·08	0·1700	0·16	−0·0954	0·45	−0·1351		

It is easily seen that at the origin $n'_o(\varepsilon, \eta) > n_o(\varepsilon, \eta)$ and hence in the immediate neighbourhood of the origin and this is the region in which we are most interested, Bernoulli's theorem gives a closer limit than Tshebycheff's analogue.

It is well known that the proof of Bernoulli's theorem (even as modified by later mathematicians is long and complicated though extremely ingenious, while Tshebycheff's theorem and its analogue rest on a proof which is as simple as it is short. Hence Bernoulli's theorem (1.1) and (1.2) might have been discarded in favour of Tshebycheff's analogue (2.4) and (2.5) if only $n'_o(\varepsilon,\eta)$ were less than $n_o(\varepsilon, \eta)$ within the region of interest in the immediate neighbourhood of the point $\varepsilon=0$, $\eta=0$, but our investigation shows that near the origin it is just the other way about, and hence the arduous mathematical processes involved in the proof of (1.1) and (1.2) are fully justified, as against the shorter and simpler reasoning behind (2.4) and (2.5).

Paper received : 1 January, 1943.

INCIDENCE OF TICKETLESS TRAVELLING ON THE LUCKNOW CAWNPORE SECTION OF THE EAST INDIAN RAILWAY

By J. SINGH

Commercial Officer, E.I. Railway

In this article it is proposed to apply the modern statistical theory of random sampling to the problem of estimating the extent of ticketless travel on a short section of the East Indian Railway, viz., the Lucknow-Cawnpore Section. As is well known the modern statistical theory of random sampling enables us under certain conditions to estimate statistical measures pertaining to a population, such as its mean, standard deviation, etc., by an analysis of a small sample drawn at random from it. Our object is to deduce the average number of ticketless passengers per day over the Lucknow-Cawnpore Section during a specified period, say 1-10-41 to 30-4-42, by examining the number of ticketless travellers detected on a small number of days selected at random during this period.

Table 1 gives the number of ticketless passengers charged by the Travelling Ticket Examiners, the total and per capita amount realised from them, the number of ticketless passengers made over for prosecution and the total number of ticketless passengers detected on the 20 random days selected.

TABLE 1. NUMBER OF TICKETLESS PASSENGERS AND THE AMOUNT REALISED FROM THEM.

date of checking	charged by examiner			number prose cuted	total number ticket-less	date of checking	charged by examiner			number prose-cuted	total number ticket-less
	number	total amount Rs.	average amount Rs.				number	total amount Rs.	average Amount Rs.		
(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
18-10-41	60	120·08	2·08	2	62	22-1-42	12	20·42	1·67	1	13
28-10-41	38	44·44	1·17	1	39	24-1-42	14	19·08	1·37	2	16
7-11-41	44	66·27	1·51	1	45	18-2-42	40	47·33	1·18	2	42
15-11-41	23	28·38	1·23	—	23	20-2-42	7	14·00	2·00	2	9
22-11-41	26	28·31	1·09	4	30	24-2-42	15	20·88	1·39	—	15
29-11-41	37	65·75	1·78	2	39	4-8-42	17	17·53	1·04	2	19
3-12-41	26	30·00	1·15	3	29	11-3-42	16	33·94	2·12	4	20
8-12-41	32	48·14	1·50	1	33	18-3-42	28	38·62	1·38	2	30
19-12-41	11	15·25	1. 39	1	12	15-4-42	34	78·47	2·31	3	37
5- 1-42	30	84·78	2·83	2	32	8-4-42	25	28·02	1·12	3	28

The first step consists in examining the sample to find out whether the population from which the sample is drawn is normal or not. For this purpose the well known statistics k_1, k_2, k_3, k_4, g_1 and g_2 introduced by Fisher are calculated. For our sample consisting of 20 items we obtain in respect of the number charged, $k_1 = 26 \cdot 75$, $k_2 = 173 \cdot 039$, $k_3 = 1551 \cdot 9956$, $k_4 = 16936 \cdot 8732$, $g_1 = 0 \cdot 68183$ and $g_2 = 0 \cdot 56564$. Standard error of $g_1 = 0 \cdot 512$ and standard error of $g_2 = 0 \cdot 9924$.

From the values of g_1 and g_2 given above, it is clear that both g_1 and g_2 are well within twice their standard errors and therefore do not deviate significantly from zero. The infinite hypothetical population may therefore be taken to be normal. Best estimates of the parameters m and σ, the mean and the standard deviation respectively for the population are provided by k_1 and $\sqrt{k_2}$ calculated above. The mean value for the sample is $k_1 = 26 \cdot 75$ and the standard error for the sample mean is $\sqrt{k_2/20} = 2 \cdot 94$. Hence the true mean m of the population must lie between $k_1 \pm 2$ (standard error), or $26 \cdot 75 \pm 2(2 \cdot 94)$. The chance of the true mean of the population lying beyond this range is less than 5 per cent and therefore, negligible. We have thus calculated the maximum and the minimum probable values of the average number of ticketless passengers travelling over the Lucknow-Cawnpore Section during the period specified above. The chances of the average number of ticketless passengers per day falling below 20·87 or exceeding 32·63 are exceedingly remote.

In a similar manner we can study the distribution of the amount realised per ticketless passenger i.e. the distribution of 'per capita realisation'. It is considered more appropriate to study the distribution of 'per capita realisation' rather than that of the total amount realised. Columns (2), (3) and (4) of Table 1 give the number of ticketless passengers charged on 20 different days, the total amount realised on these days and the amount realised per capita respectively. The values of k_1, k_2, k_3, k_4, g_1 and g_2 for the distribution of per capita realisation as calculated from the sample are respectively 1 ·56, 0 ·226, 0 ·119, 0 ·0446, 1 ·112 and 0 ·875. The standard errors of g_1 and g_2 are 0 ·512 and 0 ·873 respectively. From the values of g_1 and g_2 it is seen that g_2 is well within twice its standard error, but g_1 exceeds it by a slight amount, viz, 0 ·088. As this amount is very small, we can fairly take both g_1 and g_2 to be within twice their standard errors and thus take the infinite hypothetical population of per capita realisation also to be normal. The best estimates of the mean and variance $k_1 = 1 \cdot 56$ and $k_2 = 0 \cdot 226$. The true mean of the infinite hypothetical population of per capita realisation will here lie between $1 \cdot 56 \pm 2(0 \cdot 106)$, or between 1 ·348 and 1 ·772 in roughly 95 per cent cases.

It will thus be seen that the average number of ticketless passengers lies between 20 ·85 and 32 ·65 and the average amount realised per passenger lies between Rs. 1 ·35 and Rs. 1 ·77. Thus the average amount realised per day lies between Rs. 28 ·15 and Rs. 57 ·79.

We now proceed to estimate the average number of passengers who refuse to pay and have therefore to be prosecuted, hereinafter called the *incorrigibles*. The frequency distribution of the incorrigibles showing the number of days in which 0, 1, 2, 3 and 4 of them are detected is shown in cols. (1) and (2) of Table 2. From well-known statistical considerations, it was inferred that the population frequency distribution for this sample is Poissonian. This inference can be tested by calculating the expected frequencies on this assumption and then applying the usual (χ^2, p)-test for seeing that the actual frequencies do not differ significantly from the expected frequencies.

We now calculate the expected frequencies by substituting for the mean number of incorrigibles per day calculated from the sample, viz, 1 ·9 in $ne^{-m}\, mx/x\,!$, where $x = 0, 1, 2, \ldots$, $m = 1 \cdot 9$, $n = 20$. The expected frequencies are shown in col. (3) of Table 2. For the frequency distribution under consideration, χ^2 is seen to be 1 ·8171. The probability of χ^2 being 1 ·8171 or higher for three degrees of freedom as in the present instance, lies between 0 ·5 and 0 ·7. Thus the deviation of the observed from expected values is insignificant. The Poissonian distribution is, therefore, an appropriate fit for our data. This is also clearly shown by the close agreement of observed and expected frequencies. The average number of incorrigibles is thus seen to be 1 ·9 . The variance of the mean is $m/20$ or the standard error is 0 ·308 The chance of the true mean lying between $1 \cdot 9 \pm 2(\ 308) = 1 \cdot 9 \pm \cdot 616$ or between 1 ·284 and 2 ·516 is 95 per cent.

TABLE 2. DAYS CLASSIFIED ACCORDING TO THE NUMBER OF INCORRIGIBLES DETECTED.

no. of incorrigibles. (1)	observed frequency (2)	expected frequency (3)	$\frac{(O-E)^2}{E}$ (4)
0	2	2 ·9914	0 ·3236
1	5	5 ·6836	0 ·0822
2	8	5 ·3994	1 ·2526
3	3	3 ·4196	0 ·0515
4	2	2 ·5060	0 ·1022
Totals	20	20 ·0000	$\chi^2 = 1 \cdot 8171$ D.F. = 3

SUMMARY OF CONCLUSIONS

(1) The hypothetical infinite population of ticketless passengers from which our samples has been drawn is normal, as also the population of per capita realization.

(2) There is roughly 5 per cent chance of the true average number of ticketless passengers per day lying outside the range 20 ·85 and 32 ·65, and of the true average per capita realization lying outside the range Rs. 1 ·35 to Rs. 1 ·77.

(3) The true average amount realised from ticketless passengers per day is seen to lie in the range Rs. 28 ·15 to Rs. 57 ·79.

(4) The population of incorrigibles from which our sample has been drawn is a Poisson distribution.

(5) The chance of the true average number of incorrigibles lying between 1 ·284 and 2 ·516 is 95 per cent.

Paper received : 20 July, 1943.

INFLUENCE OF HUMIDITY AND TEMPERATURE ON THE YIELD OF COTTON

By K. RAGHAVAN NAIR and PURNENDU BOSE
Statistical Laboratory, Calcutta.

Fisher (1924) devised a special technique by which he studied the influence of weekly rainfall on the annual yield of wheat at Rothamsted. This paper is a generalisation of the above method when the influence of more than one meteorological factor is to be studied.

General formulae. Let $M_1, M_2, \ldots, M_k$ be k meteorological factors which influence the yield of the crop y. Let q be the subdivisions of the year for each of which records of the meteorological factors are available. The multiple linear regression equation of the yield of the crop upon the k meteorological factors is written as

$$Y=C+\sum_{n=1}^{q} a_{1n}m_{1n}+\sum_{n=1}^{q} a_{2n}m_{3n}+\ldots+\sum_{n=1}^{q} a_{kn}m_{kn} \quad \ldots (1)$$

where $m_{11}, m_{12}, \ldots, m_{1q}, m_{21}, m_{22}, \ldots, m_{2q}, m_{k1}, m_{k2}, \ldots, m_{kq}$ are the measurements of the meteorological factors in the different intervals of time. If the subdivisions of time were made infinitely small we should replace the linear regression function by a regression integral of the form

$$Y=C+\int_0^T a_1m_1dt+\int_0^T a_2m_2dt+\ldots+\int_0^T a_km_kdt \quad \ldots (2)$$

where m_kdt in general is the effect of the meteorological factor in the element of time dt. The integral is taken over the whole period.

If $T_0, T_1, T_2, \ldots, T_n$ be a series of orthogonal functions of time such that

$$\int_0^T T_rT_sdt=0 \quad (r\neq s), \quad \int_0^T T_r^2dt=1.$$

then we may represent the series of values of each of the meteorological factors as a time series using the following orthogonal polynomials.

$$\left.\begin{aligned} m_1&=\rho_{10}T_0+\rho_{11}T_1+\rho_{12}T_2+\ldots,+\rho_{1n}T_n \\ m_2&=\rho_{20}T_0+\rho_{21}T_1+\rho_{22}T_2+\ldots+\rho_{2n}T_n \\ m_k&=\rho_{k0}T_0+\rho_{k1}T_1+\rho_{k2}T_2+\ldots+\rho_{kn}T_n \end{aligned}\right\} \quad \ldots (3)$$

where

$$\rho_{is}=\int_0^T m_iT_sdt$$

The regression values $a_{11}, a_{12}, \ldots, a_{1n}$, for the meteorological factors in (1) may be expected to lie on a continuous time curve so that we may express the regression function in (2) in the form

$$\left.\begin{aligned} a_1&=\alpha_{10}T_0+\alpha_{11}T_1+\alpha_{12}T_2+\cdots \\ a_2&=\alpha_{20}T_0+\alpha_{21}T_1+\alpha_{22}T_2+\cdots \\ a_k&=\alpha_{k0}T_0+\alpha_{k1}T_1+\alpha_{k2}T_2+\cdots \end{aligned}\right\} \quad (4)$$

where

$$\alpha_{is}=\int_0^T a_iT_sdt$$

The degree of the polynomials in (3) and (4) should be the same and in practice seldom exceeds 4 or 5. Now using the relations (3) and (4) the relation (2) stands as

$$\begin{aligned} Y=C+(\rho_{10}\alpha_{10}+\rho_{11}\alpha_{11}+\ldots+\rho_{1n}\alpha_{1n})+(\rho_{20}\alpha_{20}+\rho_{21}\alpha_{21}+\ldots+\rho_{2n}\alpha_{2n}) \\ +\ldots+(\rho_{k0}\alpha_{k0}+\rho_{k1}\alpha_{k1}+\ldots+\rho_{kn}\alpha_{kn}) \end{aligned} \quad \ldots (5)$$

Now the values $\rho_{10},\rho_{11},\ldots,\rho_{k0},\rho_{k1},\ldots$ may be obtained for each year by fitting orthogonal polynomials for each meteorological factor, then correlating this series with the yield of the crop for several years, we can get values of α_{is} ($i=1$ to k; $s=1$ to n) as partial regression coefficients using relation (5).

By substituting these values in (4) we get values of $\alpha_1, \alpha_2, \ldots, \alpha_k$ which will give us the effect of fluctuations in each meteorological factor at any point of time in the season on the yield of the crop at the end of the season. In this paper one actual example has been worked out with two meteorological factors. In this special case equation (5) will reduce to

$$Y=C+(\rho_{10}\alpha_0+\rho_{11}\alpha_1+\ \ldots\ +\rho_{1n}\alpha_n)+(\rho_{20}\beta_0+\rho_{21}\beta_1+\ \ldots\ +\rho_{2n}\beta_n) \quad \ldots \quad (5{\cdot}1)$$

The description and the method of analysis of the experiment is discussed in the next paragraph.

Numerical Illustration. The application relates to yield of cotton recorded in the agricultural research station, Sarkand (Sind) and supplied to us for investigation by the Directof of Agriculture, Sind. In certain years there were low yields in particular tracts; and it was apprehended that such low yields were due to climatic factors. Yield figures for 1931-1941 (year 1937 was omitted) along with certain humidity and maximum temperature figures for four months (August to November) for each year were sent to us for analysis. The number of years is too small, but the material has been analysed to elucidate the practical procedure. For each day there were four readings for temperature and humidity from 10 A.M. to 4 P.M. at inter- vals of 2 hours; the mean of these four readings was taken as the representative figure for the day. We have thus 119 individual figures for temperature and humidity for each year; these were analysed by calculating the coefficients of the polynomial up to 3rd degree. The distribution of temperature and humidity in each season is thus represented by four figures.

The computation of temperature and humidity coefficients involves a great deal of labour. The method of successive summation first given by Hardy and subsequently by Fisher was used. The method may be briefly described as follows.

Let the whole interval be divided into two parts (1 to m) and (m to n). Then,

$$s_1=\sum_{i=1}^{m} x_1,\ s_2=\sum_{i=1}^{m}\left\{(m+1-i)\,x_1\right\},\ s_3=\sum_{i=1}^{m}\left\{(m+1-i)\,(m+2-i)\,x_1/2!\right\}$$

$$s_k=\sum_{i=1}^{m}\left\{(m+1-i)\,(m+2-i)\ldots(m+k-1-i)x_1/(k-1)!\right\}$$

where x_1 are the numbers. The remaining n numbers are to be summed backwards from the bottom of the column, dropping one term at the end of each summation and so giving

$$s_1'=\sum_{i=m}^{n} x_1,\ s_2'=\sum_{i=m}^{n}\left\{(m+1-i)\,x_1\right\},\ s_3'=\sum_{i=m}^{n}\left\{(m+1-i)(m+2-i)x_1/2!\right\}$$

$$s_k'=\sum_{i=m}^{n}\left\{(m+1-i)\,(m+2-i)\ \ldots\ (m+k-1-i)\,x_1/(k-1)!\right\}$$

Taking alternately sums and differences we obtain

$$S_1'=s_1+s_1',\quad S_2'=s_2-s_2',\quad S_3'=s_3+s_3',\quad S_4'=s_4-s_4'$$

and so on. from which the final sums of the whole column may be obtained from the equation

$$S_1=S_1',\quad S_2=S_2'+MS_1',\quad S_3=S_3'+MS_2'+M\,(M+1)\,S_1'/2!$$

$$S_4=S_4'+MS_3'+M\,(M+1)\,S_2'/2!+M(M+1)\,(M+2)\,S_3'/3!$$

The series $S_1, S_2, S_3, S_4, \ldots$ has to be divided by a series of numbers $n,\ n\,(n+1)/2!,\ n(n+1)\,(n+2)/3!,\ n(n+1)(n+2)(n+3)/4!$ etc. yielding a series of numbers.

$$a=\frac{1}{n}\,\Sigma(x_1),\quad b=\frac{2!}{n(n+1)}\Sigma\{(n+1-i)x_1\},\quad c=\frac{3!}{n(n+1)(n+2)}\Sigma\{(n+1-i)\,(n+2-i)x_1/2!\}$$

$$d=\frac{4!}{n(n+1)(n+2)(n+3)}\,\Sigma\{(n+1-i)\,(n+2-i)\,(n+3-i)\,x_1/3!\}$$

In the present case as we have fitted a third degree curve, it is sufficient to calculate up to c. From the values of $a, b, c, d, \ldots$ we can finally calculate ρ-values from the following relations

$$\rho_0=\sqrt{n}.a,\quad \rho_1=(a-b)\sqrt{3n\,(n+1)/(n-1)},\quad \rho_2=(a-3b+2c)\sqrt{5n(n+1)\,(n+2)/(n-1)(n-2)}$$

$$\rho_3=(a-6b+10c-5d)\sqrt{7n(n+1)\,(n+2)\,(n+3)/(n-1)(n-2)(n-3)}$$

Polynomial values of Meteorological Factors. In table I. columns (4) and (5) gives the values of S_r for different values of *r*, and the last two columns give the values of the constants *a*, *b*, *c*, and *d*, for each year.

TABLE 1. WEIGHTED MEANS OF THE SERIES OF READINGS FOR HUMIDITY AND TEMPERATURE.

year	constant	r	S_r		values of the constant	
			humidity	temperature	humidity	temperature
(1)	(2)	(3)	(4)	(5)	(5)	(7)
1931	a	1	6547 ·4	10569 ·8	55 ·0202	88 ·8218
	b	2	425440 ·7	649440 ·8	59 ·5883	90 ·9581
	c	3	17665140 ·7	26517915 ·2	61 ·3416	92 ·0825
	d	4	544473478 ·2	814500515 ·4	61 ·9890	92 ·7319
1932	a	1	5540 ·7	10581 ·2	46 ·5605	88 ·9176
	b	2	378876 ·9	656769 ·8	53 ·0640	91 ·9846
	c	3	16332976 ·5	26780306 ·0	56 ·7157	92 ·9936
	d	4	517979374 ·9	819840926 ·8	58 ·9726	93 ·3399
1933	a	1	6060 ·5	10650 ·47	50 ·9286	89 ·4997
	b	2	390111 ·7	654479 ·41	54 ·6375	91 ·6638
	c	3	16403599 ·4	26514585 ·22	56 ·9609	92 ·0709
	d	3	512533025 ·8	809598382 ·30	58 ·3525	92 ·1738
1934	a	1	5138 ·7	10006 ·0	43 ·1824	84 ·0840
	b	2	346501 ·4	632339 ·4	48 ·5296	88 ·5629
	c	3	14742044 ·9	26017704 ·8	51 ·1912	90 ·3455
	d	4	460845528 ·0	800895752 ·2	52 ·4678	91 ·1830
1935	a	1	4565 ·7	10289 ·1	38 ·3672	86 ·4630
	b	2	287612 ·0	643466 ·9	40 ·2818	80 ·1214
	c	3	11984684 ·4	25402355 ·8	41 ·6164	91 ·6812
	d	4	371828888 ·6	811902273 ·8	42 ·3326	92 ·4361
1936	a	1	4122 ·1	10518 ·0	34 ·6395	88 ·3866
	b	2	257847 ·0	657457 ·8	36 ·1130	92 ·0809
	c	3	10756772 ·3	26882440 ·4	37 ·3525	93 ·3483
	d	4	334795700 ·9	824728215 · ·6	38 ·1169	93 ·8963
1938	a	1	4446 ·4	11107 ·6	37 ·3647	93 ·3412
	b	2	306766 ·6	692000 ·4	42 ·9645	96 ·9188
	c	3	13197181 ·0	28179982 ·2	45 ·8267	97 ·8540
	d	4	415306927 ·8	862326901 ·9	47 ·2809	98 ·1770
1939	a	1	3816 ·0	11277 ·6	32 ·0672	94 ·7697
	b	2	259626 ·9	697233 ·2	36 ·3623	97 ·6517
	c	3	11129474 ·7	28351035 ·0	38 ·6467	98 ·4479
	d	4	350337184 ·4	867277874 ·0	39 ·8863	98 ·7407
1940	a	1	4167 ·5	11340 ·7	35 ·0210	95 ·3000
	b	2	279030 ·4	695126 ·2	39 ·0799	97 ·3566
	c	3	12002542 ·0	28168959 ·5	41 ·6784	97 ·8157
	d	4	379086284 ·4	861296903 ·6	43 ·1479	98 ·0597

Table 2 gives the orthogonal polynomial coefficients for humidity and temperature for each year in columns (3) to (8). Col. (2) gives the yield of cotton for each year.

Regression Equations. Having obtained the ρ values for temperature and humidity we have fitted separately (1) yield against humidity ; (2) yield against temperature ; and (3) yield jointly against humidity and temperature. The respective equations are as follows .. (i)

$$Y=112 \cdot 8274+2 \cdot 1910\, \rho_{10}-0 \cdot 7752\, \rho_{11}+4 \cdot 1904\, \rho_{12} \quad .. \text{(ii)}$$

$$Y=1613 \cdot 4737+0 \cdot 0002\, \rho_{20}-0 \cdot 0122\, \rho_{21}+16 \cdot 0435\, \rho_{22}$$

$$Y=493 \cdot 0320+2 \cdot 1283\, \rho_{10}-0 \cdot 0762\, \rho_{11}+3 \cdot 5696\, \rho_{12}-0 \cdot 001584\, \rho_{20}-0 \cdot 0050\, \rho_{21}+11 \cdot 4488\, \rho_{22} \quad .. \text{(iii)}$$

Values of the multiple correlation coefficients are given below.

$$R_1=0 \cdot 3079, \quad R_2=0 \cdot 5099, \quad R_3=0 \cdot 5646$$

The degrees of freedom for testing R_3, R_2, and R_1 are 5, 5 and 2 respectively, but none of them are significant.

Table 2. Orthogonal-polynomial coefficients for humidity and temperature

year	yield	humidity			temperature		
		ρ_{10}	ρ_{11}	ρ_{12}	ρ_{20}	ρ_{21}	ρ_{22}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1931	1694·2	600·27	—87·20	—26·56	969·05	—40·70	—2·81
1932	591·5	507·98	—123·89	20·01	970·09	—58·43	—26·25
1933	1701·1	555·63	—70·65	23·47	976·44	—41·23	—33·77
1934	1328·0	471·12	—101·88	—0·65	917·36	—85·32	—22·86
1935	1414·0	418·59	—36·47	18·88	943·31	—69·69	—13·48
1936	1389·0	377·91	—28·08	25·11	964·30	—70·38	—29·01
1938	655·0	407·60	—106·68	3·50	1018·35	—68·15	—42·71
1939	1274·0	349·85	—81·82	6·85	1033·94	—54·90	—32·27
1940	857·0	382·07	—77·34	28·52	1039·72	—39·20	—28·48

Conclusion. On available evidence there is no significant influence of temperature and humidity on the yield of cotton. It must however be remembered that owing to the scantiness of the material analysed here, any but a large influence might well remain undetected.

REFERENCES

1. Fisher R. A. (1924) The influence of rainfall on the yield of wheat at Rothamsted. *Philoophical Transactions of the Royal Society of London*, B, **213**, 89-142.

Paper received : 20 *March*, 1943.

INDIAN STATISTICAL INSTITUTE

Thirteenth Annual Report: 1944-45

The Indian Statistical Institute, which was inaugurated on 17 December 1931 and was formally constituted into a non-profit making society registered under Act XXI of 1860 on 28 April 1932, has completed thirteen years of useful work. Its activities this year were even more varied than in the past thus reflecting the rapid growth of the application of statistical methods to problems of human welfare in India. The Institute combines in one organization a scientific society, a centre for advanced training and researches in statistics, and a fact finding agency using the technique of sample surveys on a large scale.

The total number of worker-months employed in the statistical branch during the year was 5220 with an average of 435 workers per month; and in the field branch, 5000 with an average of 417 workers per month. The statistical work was done as before at three places—in Calcutta, at Baranagore and at Giridih. Owing to limited accommodation in Calcutta, the major portion of the staff was posted at Baranagore and Giridih, the actual number of workers in March 1945 being 161 and 178 at these two places respectively.

CROP SURVEY WORK IN BENGAL

The sample survey of crop acreage and outturn entered in its eighth year of progress in 1944-45. The Government of Bengal had originally sanctioned a consolidated grant of Rs. 4,20,000 for carrying out a two-season sample survey of the jute, *aus* paddy, and *aman* paddy crops in Bengal in 1944-45. In May 1944, H. E. Mr. Casey, Governor of Bengal, sent for P. C. Mahalanobis, and after personal discussions decided to sanction a three year sample survey programme on an enlarged scale for which the Government of Bengal agreed to pay a lump grant of Rs. 7, 17, 000 in the first year (1944-45) and Rs. 6,00,000 in each of the two subsequent years. Government also agreed to pay dearness allowance to the staff engaged in this work according to rates prevalent from time to time in Government service. The proposal was accepted by the Institute. The preparatory work of the jute-*aus* season had already been completed by this time and the field work also had made considerable progress. Work was, however, taken in hand immediately for giving effect to this enlarged scheme from the *aman* paddy season of 1944-45.

Design of the Survey : The whole province was divided into zones each approximately 64 square miles in area. In the *aman* season each of these zones was divided into 64 cells (of one sq. mile each in area) and out of these 26 cells were selected at random in the first instance. In each cell so selected 4 grids of the uniform size of 2.25 acres were located at random, thus securing an average density of 1.6 grid per sq. mile for area enumeration. Crop-cutting was done in 240 zones selected at random, and in 16 grids selected at random out of 104 in each zone. The survey was arranged (as in previous years) in the form of two inter-penetrating net-works of grids for each of which the information was collected independently by different parties of field investigators. About 8% of the grids were common to both net-works and were surveyed in duplicate by both parties.

Field organization : Mr. N. C. Chakravarti, M.A., B.C.S., Additional Assistant Secretary in the Agriculture Department, whose part-time services had been lent by Government for conducting the field survey under the Institute, continued to be in charge of the work throughout the year. The field staff under him consisted on an average of 4 supervising officers, 16 chief inspectors, 64 inspectors and 316 field workers.

Progress of Survey : The area survey of jute and *aus* paddy was started in East Bengal districts by the middle of May 1944 and was extended to other districts in the course of a few weeks. The bulk of the area survey and crop-cutting work on jute and *aus* paddy was completed by the end of August 1944, but contiuued in certain areas till October 1944. The *aman* survey started from the beginning of September 1944 and was completed by the end of January 1945. A total of over 64,000 grids had been surveyed in the jute-*aus* season, together with an additional batch of about 5,000 grids which were surveyed for a second time during the harvest period. In the *aman* season about 1,05,000 grids were surveyed at first followed by a second survey during harvest of over 8000 grids.

Analysis of data : Tabulation of field records was arranged in such a way as to supply estimates of acreage and of outturn based on the material collected upto a fortnight previous to the date of report. This enabled progressive estimates being submitted to the Government every fortnight.

It may be mentioned here that along with the main area survey and crop-cutting experiments, various investigations were carried out for improving the technique of the sample survey. The total production of the *aman* paddy crop, for example, was estimated directly from the estimated production of the individual grids surveyed at harvest time. Different sizes and shapes of cuts as well as different types of cutting apparatus were used in crop-cutting work; and experiments were also made to explore the possibilities of rapid eye-estimating of rates of yield. An account of the technical work will be given in scientific reports.

Along with the sample survey in 1944-45, the Government of Bengal had arranged for a simultaneous "complete enumeration" of crops on all plots being carried out by the Development Commissioner with the help of a large field-staff consisting of about 18,000 field workers and 1500 inspectors and supervising officers at a cost, it is understood, of nearly Rs. 50,00,000. The results of the complete enumeration have not, however, been sent to the Institute so far. It is not possible, therefore, to compare these results with those of the sample survey.

At the request of the Government of Bengal, Prof. R. A. Fisher during his recent visit to India examined both the schemes, the one for sample survey as well as that for complete enumeration. His opinion is quoted below :—

> "The sample survey of acreage as at present actually conducted in Bengal and Bihar could be depended upon to give provincial or district estimates with sufficient accuracy for all practical purposes."

Experience in the United States of America corroborates this. In an investigation of the Bureau of Agricultural Economics it was found that "in some items [including crop acreage] the sample survey was more accurate than the [Federal Agricultural] Census". (*Iowa Research Bulletin* No. 304, 1942, p.94). In a later report it is stated that U.S. Census figures "are not known within ±5 per cent of the actual", and also that in a sample survey "resulting accuracies....will probably exceed that of the Census". (*Iowa Reasearch Bulletin* No. 309, 1944, p. & p. 332).

Preparatory work for 1945-46 : The preparatory work for the 1945-46 survey has been taken up early in 1945 and has made satisfactory progress.

Crop Survey Work in Bihar

Previous history : As a result of discussions with His Excellency Mr. (now Sir) Francis Mudie, then Governor of Bihar, a five-year scheme for the improvement of crop statistics in Bihar by the method of sample surveys had been sanctioned in 1943. An exploratory survey of *rabi* crops in two districts Monghyr and Shahabad was started on 12 February and completed on 13 April 1944. A large number of field workers recruited in Bihar were trained up by a contingent of about 80 experienced field workers who were sent from Bengal to Bihar for this purpose. Preliminary reports were sent from time to time and the final report was submitted in November 1944.

The work in the exploratory survey was organized with such great success that it was decided to take up full scale provincial surveys from the *bhadoi* season of 1944. The field work in this scheme was done under the direct control of the Bihar Government. The Institute was responsible for the whole of the statistical work, for which a block grant of Rs. 20,000 per month was sanctioned for the period April 1944—June 1945. The bulk of the statistical work was done at Giridih.

Work in 1944-45 : In the current financial year sample surveys covering the whole province were carried out in three different seasons: *bhadoi*, *aghani*, and *rabi* 1944-45 ; a survey of the early *makai* crop was also undertaken in 3 districts.

Design of the Survey : In *bhadoi* and subsequent surveys the whole province was divided into 1343 cells, each of approximately 50 square miles in area and roughly square in shape. In each season the cells were classified according to the intensity level of the crops under survey and were grouped into a suitable number of blocks of more or less homogeneous intensity. A suitable number of grids (or sample-units of size 4-acre each) were located at random in the different blocks, and were surveyed by field investigators. The whole work was arranged in the form of two interpenetrating net-works of samples information for each of which was collected independently by a different party of investigators. Certain grids were also surveyed in duplicate by both the parties and supplied a check on the accuracy of the field work. The total number of grids used in the three provincial surveys varied from about 65,000 to 78,000. Crop-cutting work on paddy, wheat, gram and *masur* was done in a suitable number of randomly selected grids in which the area survey was also repeated at the time of harvesting. This supplied material for estimating the yield per acre of each crop as well as the total outturn directly from the joint area and yield survey

Field organization : Rai Bahadur S. P. Sinha, a senior Deputy Collector, was throughout in charge of the field work. In the *rabi* survey of 1943-44 a strong batch of Bengal field workers conducted the field survey and gave training to the newly recruited Bihar workers. The Bihar workers were responsible for the entire field work in all the other surveys. Each field unit consisted of 4 investigators in charge of 1 inspector; and 4 such units were placed under one chief inspector. There were 64 field units with 16 chief inspectors, 64 inspectors and 256 investigators for the whole province. Besides this a number of investigators and inspectors were engaged in special work.

Progress of survey : The early *makai* survey in portions of districts of Monghyr, Bhagalpur and Purnia was carried out in May and June 1944. The *bhadoi* survey (which was the first survey on a provincial scale) started on 3 July 1944 and was completed on 6 October 1944. The *aghani* survey started on 10 October 1944 almost immediately after the termination of the *bhadoi* survey and continued till the middle of January 1945. This was followed by the *rabi* survey which started in the middle of January and was completed on 10 April 1945.

Statistical work : As in the Bengal scheme, tabulation of field data was arranged in such a way as to supply crop estimates at intervals of from 10 to 15 days in the form of progressive reports. The preliminary analysis of the *makai* survey was finished during the progress of the *bhadoi* survey. A report of *bhadoi* survey based on full material was submitted early in November 1944; eight progress reports on the *aghani* survey were submitted from time to time and a report based on the full material was submitted in March 1945; and four progress reports have already been submitted on the *rabi* survey.

Special studies : Apart from the general survey of acreage and outturn various special studies were made in all the surveys. P. C. Mahalanobis visited a special study camp at Bikramganj during the *rabi* season of 1943-44. In the *bhadoi* season studies on cost and variance function for area survey were made at Giridih; and large scale studies were undertaken on "complete enumeration" of crops. In the *aghani* season apart from cost and variance studies special investigations on the technique of crop-cutting work were undertaken at Giridih, Gaya and Patna; similar work was undertaken in the *rabi* season at Patna, Gaya and Pusa. Whenever possible such special study camps were placed under the charge of qualified Institute workers with knowledge of statistical theory.

Decision to terminate the sample survey scheme in Bihar : We regret to report that without any previous discussion with the Institute the Government of Bihar decided to terminate the sample survey scheme and substitute in its place a scheme of "complete enumeration" of crops. As already noted, Prof. R. A. Fisher, F.R.S. expressed the opinion that the sample survey as actually carried out during the last 6 months in Bihar could be depended upon to supply crop estimates with a precision adequate for all practical purposes. In a permanently settled area like Bihar, lacking as it does a well-established *patwari* system, it is not considered likely that accurate estimates of crop acreage can be obtained in the immediate future by the method of complete enumeration. Discontinuance of the sample survey would, therefore, leave a serious gap in crop statistics, and may in a bad year lead to the same dangerous situation as prevailed in Bengal in 1943. In written notes submitted to the Government of India and in the course of various discussions and conferences at New Delhi both Professors R. A. Fisher, F.R.S. and P. C. Mahalanobis, F.R.S. have strongly urged the continuance of the sample survey scheme in Bihar.

POPULATION DATA COMMITTEE

In March 1944 P. C. Mahalanobis went to Delhi at the request of the Department of Education, Health and Lands, to discuss possibilities of estimating the growth of population in India in future. As a result of informal discussions the Government of India in their order No. D.2249-H(c)/44 dated 15 May 1944, constituted "a small expert committee to examine and advise the Government of India on the available data relating to growth of population" with Mr. M. W. M. Yeatts, I.C.S., (Census Commissioner in 1941) as Chairman, Sir Theodre Gregory, Dr. K. C. K. E. Raja, and Professors K. B. Madhava and P. C. Mahalanobis as members. The work of the Committee was based at the Indian Statistical Institute and a grant of Rs. 54,000 to the Institute was sanctioned for this purpose. The services of Prof. Madhava were secured on loan from the Government of Mysore and were placed at the disposal of the Institute by the Government of India for this work.

As a measure of retrenchment most of the census tabulation work (including the preparation of age-tables) had been cut down in 1941. Fortunately, Mr. Yeatts (as Census Commissioner) had issued instructions for extracting and preserving every fiftieth individual slip in each census area. This led to the preservation of a two per cent sample (called Y-sample) for the bulk of the original census material. The sample slips numbering about 7 millions were brought over to the Indian Statistical Institute and appropriate statistical analysis was used to test the representativeness of the Y-sample and to ascertain the precision which could be attained in reconstructing age and other tables from the Y-sample. Two interim reports were submitted to the Government of India on 11 September 1944 and 15 November 1944 respectively, and the final report is nearly ready.

The results of the investigations are on the whole satisfactory. In many areas the Y-samples behave practically as true random samples; in certain areas there were, however, significant differences between results based on complete counts and those based on the Y-sample. Such discrepancies (which might have arisen from the Y-sample not having been drawn in strict accordance with specifications) were, however, invariably small in magnitude so that age and other tables reconstructed from the Y-samples were likely to be adequate for the preparation of life tables and population projections. It is expected that the work of preparing age and life-tables and population projections would be started in the near future. If the present enterprise proves successful it will extend the scope of the sampling method very greatly in census work in future. A proposal for transferring to Hollerith cards the information on the available sample slips was submitted to the Government of India by the Institute in November 1944 and is under consideration.

Bengal Distressed Villages Enquiry 1944-45

This survey was taken up at the instance of the Government of Bengal with a view to estimating the extent of damage done to rural economy by famine conditions in Bengal in 1943, and a block grant of Rs. 25,000 was sanctioned by Government for this purpose. The field survey was organized with the active collaboration and under the general control of Prof. K. P. Chattopadhyaya of the Anthropology Department of the University of Calcutta. The field work started in June 1944 and continued till the end of March 1945. Information was collected by direct enquiry by investigators for about 16,000 families selected at random from 400 villages which were also picked up at random in 40 subdivisions. Provisional reports were submitted to the Government of Bengal in February 1945, aud other reports on subsequent dates.

A survey was taken up in Jagaddal (Bhatpara) as a part of the present enquiry with a view to making a comparative study regarding economic changes in rural and urban areas. The field work was started in December 1944 and was completed by 5 March 1945. Information was collected for 761 families, and the analysis of the material is proceeding.

Enquiry into the Economics and Statistics of Road Development

In November 1944 P.C. Mahalanobis went to Delhi at the request of the War Transport Department; and, after discussing with road engineers what could be done to investigate economic and statistical aspects of road develoment, he submitted a scheme for an exploratory survey. In their order No. PL.2(5) dated 5 January 1945, Government of India (War Transport Department) sanctioned a block grant of Rs. 85,000 to the Indian Statistical Institute for the purpose, and the period of loan of Prof. K. B. Madhava's services from the Mysore Government was extended till the end of 1945 for this scheme.

Arrangements were made to conduct a preliminary enquiry in several places in Bengal and Bihar, and by the end of March information was collected for 18 villages. In the original scheme it had been decided that the Government of India would appoint an Engineer whose services would be made available for this enquiry. This appointment had not however been made even by the end of March 1945, which prevented traffic census being completed in these areas, and hampered progress.

Other Projects and Special Enquiries

A large number of other projects and special enquiries were also taken up.

Burdwan-Howrah-Hooghly Irrigation Scheme : Crop-cutting work on *aman* paddy had been carried out for six years (1938-39 to 1943-44) in this scheme to assess and grade the different classes of soil in terms of fertility as measured by rates of yield per acre of paddy. But after all this work had been done a scrutiny in the Statistical Laboratory of records of soil classification revealed large discrepancies. It is now proposed to check by direct re-survey and rectify the mistakes in soil-classifications of all plots in which crop-cutting experiments were ever conducted. This survey will have to cover about 70,000 plots in 1700 villages scattered over the three districts Burdwan, Howrah and Hooghly.

U. P. Anthropometric survey : The work on the anthropometrical material collected by Dr. D. N. Majumdar in 1941 for 23 castes in the United Provinces has been practically completed. The values of Mahalanobis's D- square were calculated, and a systematic classification of castes has been prepared and a report is being got ready.

Preparation of roster of scientific and technical personnel : At the instance of the Government of India and the National Institute of Sciences some preliminary work has been done on the preparation of a roster of scientific and technical personnel in India on the lines of the roster prepared for U.K. by the Royal Society of London. Our thanks are due to Prof. A. V. Hill, F. R. S., Secretary, Royal Society, for sending us a complete set of forms and cards used in the U. K. roster.

River Research Institute : Statistical analysis of data relating to river physics, rainfall and flood was continued this year with the help of a grant of Rs. 5,000 from the River Research Institute, Bengal.

Famine Inquiry Commission : P. C. Mahalanobis submitted a written note, and also gave evidence before the Bengal Famine Inquiry Commission on 13 September 1944. At the request of the Commission the Institute prepared several notes on per capita consumption of rice and other selected food items and also on the number of deaths in Bengal in 1943 based on information collected by sample surveys. A grant of Rs. 2,000 was received from the Commission for this purpose.

Sriniketan Rural Survey : The analysis of material relating to rural families collected by the Visva-bharati Institute of Rural Reconstruction at Sriniketan has been completed at Giridih and a paper on the same has been prepared.

MISCELLANEOUS STATISTICAL ENQUIRIES

During the year under review, a number of miscellaneous statistical enquiries were attended to as usual. The distribution by subjects and administrative regions of important items are shown below :

By subjects : Agriculture 21 ; Educational 3 ; Industrial 8 ; Meteorological 2 ; Sample Survey 2 ; Socio-economic 7 ; Theoretical 3 ; Vital and Public Health 3 ; Other subjects 11 ; Total 60.

By regions : Assam 1 ; Bengal 22 ; Bihar 4 ; Bombay 2 ; C.P. and Berar 1 ; Ceylon 1 ; Delhi 10 ; Orissa 6 ; Punjab 1 ; Sind 2 ; Travancore 2 ; U. P. 8 ; Total 60.

THEORETICAL RESEARCHES

Theoretical work was continued on the following main topics : (1) design of large scale sample surveys ; (2) design and analysis of experiments ; (3) sampling distributions ; (4) testing of hypothesis ; and (5) construction of numerical tables. A list of papers published and reports submitted during the year by workers associated with the Institute is given in Appendix 1.

Large scale sample surveys : P. C. Mahalanobis's memoir on "Large Scale Sample Surveys" was published in the *Philosophical Transactions* of the Royal Society of London in October 1944. Further work is proceeding on problems relating to zoning and mapping in the design of sample survey, properties of random and non-random fields, and double sampling.

Design and analysis of experiments : Work on the design of experiments has been continued during the year by Raj Chandra Bose, C. Radhakrishna Rao, Hari Kinkar Nandi, Kailas Nath Bhattacharya and Quazi M. Hossain. Elegant solutions were obtained for some of the combinatorial problems remaining unsolved in the construction of balanced incomplete block designs. Some new cyclic solutions have been given and the impossibility of some of the combinatorial problems has been demonstrated. A new line of study as to the enumeration of the number of non-isomorphic solutions to the incomplete block design has been opened and much progress has been made. A rigorous proof as to the conditions under which Youden's squares exist and can be constructed has been advanced. A study has been made as to the inequality relations connecting the number of blocks, treatments, plots per block etc. for a general incomplete block design. Further work has been done on the linear hypothesis leading to the inter-block and intra-block estimation in an incomplete block design.

Sampling distributions : Samarendra Nath Roy continued his earlier work on the p-statistics, developing methods by which k multivariate normal populations (supposed to have the same dispersion matrix) differing in the set of mean values could be suitably classified by ordering them against some sort of scale. Account was taken of the essentially vectorial nature of the problem involving as it does a number of characters. Manindra Nath Ghose made some study of the order of approximation of the central limit theorem.

Testing of hypothesis : Study of (1) the problem of testing composite hypothesis, (2) the exact nature of the usual tests in analysis of variance, D^2, etc., and (3) their optimum properties, was taken up during the year by S. N. Roy, M. N. Ghose and H. K. Nandi ; and many useful results have been obtained.

Construction of tables and other problems : The construction of 5% and 1% points of the classical D^2-statistic has been completed for different values of the parameters ; the fiducial limits are now being calculated. A general study of the limiting forms of the statistical distributions (which ultimately tend to normal) has been made, and a suitable curve along with a table for different levels of error has been constructed. The special form of Bessel function (I_0 and I_1) has been studied for large values of the argument. Further work on the unbiassed partitioning of tail ends is being continued by Purnendu Kumar Bose.

Sankhya : The Indian Journal of Statistics

Printing of Sankhyā was delayed by the promulgation of the Paper Control Order by the Government of India in June 1944 which required the number of pages of every journal being reduced by 70%. As a result of representations made by the Honorary Secretary, special sanction has, however, been given by the Government of India relaxing the above limit to a considerable extent. Publication was also hampered by shortage of paper, lack of printing facilities, and other difficulties most of which have been overcome and it is expected that Vol. 7, part 1 would be published at an early date and would be soon followed by other issues.

Sankhya : Receipts & Payments Account for the Year ending **31**st March 1945.

RECEIPTS			PAYMENTS		
To Opening Balance ..			By Press Bills & Paper etc.		1,969 6 0
Cash in hand ..	2 5 6		" Closing Balance		
Cash at Bank	390 15 6	393 5 0	Cash in hand	712 5 6	
" Subscription		1,358 14 0	Cash in bank	1,320 7 6	2,032 13 0
" Grant from the Govt. of India		2,000 0 0			
" Rockefeller Foundation Grant		250 0 0			
		4,002 3 0			4,002 3 0

Training and Examinations in Statistics

Although two other centres of training in statistics are now functioning in Calcutta, viz. the four-year old post-graduate department of Statistics of the Calcutta University and the recently opened honours degree course in statistics in the Presidency College, Calcutta, we received a large number of applications for admission to the training section of the Institute. We, however, could admit only 11 because of our limited capacity for arranging practical work. This year the Bihar Government granted three scholarships and the Orissa Government one to students nominated by them for receiving training in the Institute. The training section was in charge of Birendra Nath Ghose assisted by Nirmal Kumar Chakravarty, but almost all the senior workers of the Institute took part in teaching.

Besides regular students, a large number of officers from various institutions came to the Institute for consultation ; and the heavy burden of supervising this work was borne by senior workers among whom Purnendu Kumar Bose deserves special mention.

Computers' training : A short course of training for helping the workers who appeared at the Computer's Certificate examination was given in Calcutta in the months of August, September and October 1944 under the supervision of Jitendra Mohan Sen Gupta and Sambhu Nath Halder.

Institute Examinations in Statistics : This year Computer's Certificate examinations were held once in July 1944 at Giridih and again in January 1945 both in Calcutta and Giridih. A large number of candidates appeared in different sections of the Computer's Certificate examination while only one sat for the Statistician's Diploma. Out of a total of 91 candidates 38 came out successful. The Board of Examiners consisted of Prof. P. C. Mahalanobis, Prof. K. B. Madhava, Dr. C. Chandra Sekar, and Messrs. R. C. Bose, S. N. Roy, N. T. Mathew and J. M. Sengupta.

Arrangements in the Statistical Laboratory

As noted earlier, the Statistical Laboratory had to function at three places, namely, Calcutta Baranagore and Giridih. In Calcutta itself work was carried on at 4 places—in the Presidency College, Hindu School, 210 Cornwallis Street, and a room in the Albert Institute. The Honorary Secretary went to Giridih several times during the year, and other workers were also sent there from time to time. Owing to postal delays and difficulties of communication arrangements had to be made for important records being carried between Calcutta and Giridih by special messengers.

A thorough reorganization was made of all Institute posts in a number of graded scales with effect from August 1944.

Number of workers : The number of workers employed on a monthly basis was 179 and the number of piece-rate workers 140 on 1 April 1944. During the year under review 178 new appointments were made on a monthly basis and 296 as piece-rate workers, while 78 on monthly basis and 189 piece-rate workers left the Laboratory, so that on 31 March 1945 the year closed with 279 workers on a monthly basis and 247 piece-rate workers with a total strength of 526 workers in all. The average strength of workers was 144 in Calcutta, 113 a Baranagore and 178 at Giridih during the year under review.

Institute staff : Jitendra Mohan Sen Gupta, C. Radhakrishna Rao, Nirmal Kumar Chakravarty, Manindra Nath Ghose, R. P. Bhargava, Ramkrishna Mukherjee and Mrs. Chameli Bose worked mostly in Calcutta and N. T. Mathew, K.C Cheriyan and Monimohan Mukherjee mostly at Giridih. Raj Chandra Bose, Samarendra Nath Roy, Birendranath Ghose, Purnendu Kumar Bose and Hari Kinkar Nandi, whole-time lecturers either in the Calcutta University or in the Presidency College, Calcutta, participated in the activities of the Institute as part-time workers. In the computing section, Jitendra Nath Taluqdar and Sambhu Nath Halder assisted by Arun Kumar Ganguly, Suresh Chandra Sarkhel, Shyamsundar Bose, Sunith Krishna Gupta and Nirmal Banerjee were in general charge of the work at Baranagore and Calcutta while Haribhajan Choudhury assisted by Dhirendra Nath Sarkar, Rajen Roy and others looked after the computing work at Giridih. Among old workers Sudhir Kumar Banerjee, Heramba Chandra Ghosal, Subodh Chandra Das Gupta, Jibananda Saha, Jaladhar Sarma, Naresh Chandra Mukerjee, Susanta Ranjan Guha, Gayanath Adak, Sarat Chandra Mitra, Probhat Sinha, Debabrata Dutta, Sukumar Hazra, Purna Chandra Ghose looked after various sections in Calcutta, Baranagore and Giridih.

In the field branch of the Bengal Crop Survey Scheme the part-time services of Nihar Chandra Chakravarti, B.C.S., Additional Assistant Secretary to the Government of Bengal, Department of Agriculture, were made available to the Institute. He worked as Supervisor of the Field Branch assisted by Dhirendra Mohan Ganguly, Santipriya Bose, Pranay Kumar Chatterjee and Saroj Kumar Chakravorty.

The services of Prof. K. B. Madhava of Mysore University were made available to the Institute by the Government of India from June 1944. He took a large share of the superior administrative responsibility specially during the period June-August 1944 when P. C. Mahalanobis was suffering from an attack of appendicitis. K. N. Chakravarti and H. C. Ghosh acted as part-time Bursars during the year under review.

Our best thanks are due to the University of Lucknow for placing the services of Dr. D. N. Majumdar from February 1945 at the disposal of the Institute for organizing an anthropometric survey in Bengal.

Changes in staff : After a long connection of more than 6 years S. Raja Rao left us in September 1944 to join the Public Works Department, Sind, Karachi as Statistical Officer. Nimai Charan Ghose, who was with us for more than 4 years, left us in June 1944 to join the Dept. of Central Excise, Calcutta. Manindra Nath Ghose who had left to join a temporary post of lecturer of statistics in the Dacca University rejoined us in January, 1944. Miss Bina Sen came back in March 1945 on the expiry of service as temporary lecturer of mathematics in the Lady Brabourne College, Calcutta. Anil Kumar Gayen rejoined us in December 1944 at the termination of his services as temporary lecturer in the Presidency College, Calcutta. Among the students of the Statistics Department of the Calcutta University who took their M.A. or M.Sc. degree in 1944, the following have joined the Institute in various capacities: Prosad Kumar Banerjee, Tarapada Choudhury, Durgadas Bose, Premranjan Sengupta and Prafulla Kumar Bhowmik.

Women's Section in Calcutta : One section with an average of 27 women workers employed on the piece-rate system of payment was run throughout the year under the supervision of Mrs. P. Sen at 210/6 Cornwallis Street, and a second section in charge of Mrs. Ila Sinha was opened at 180/A Upper Circular Road in February 1945.

BRANCH LABORATORY AT BARANAGORE

As already stated, owing to limited accommodation at Calcutta, a large portion of the statistical staff was housed at Baranagore. A three storied building has been constructed to accommodate the workers the number of which is increasing steadily ; the total number of workers was 161 at the end of March 1945. A separate office with a library section has been organized there. The bulk of the Bengal Crop Survey work is being done at Barangore together with portions of other project work. The Baranagore Branch had to run 3 messes near the office. A medical welfare scheme was started in February 1945 and a tiffin club for workers in March 1945.

Land acquisition : Since April 1942 the Institute has been in occupation, free of rent, premises 'Amrapali' at 87 Barrackpur Trunk Road belonging to the Honorary Secretary. Owing to expansion of work there was urgent need of additional accommodation. It was, therefore, decided to acquire about 3 acres of land adjoining 'Amrapali', for which notification under the Land Acquisition Act was issued on the

11th January 1945, and it is hoped that possession would be obtained at an early date. The cost of acquisition would be met by an advance, with the consent of P. C. Mahalanobis, from the Director's Contribution Fund.

Branch Laboratory at Giridih

The Giridih Branch of the Indian Statistical Institute had started mainly as an evacuation measure. On the inception of the Bihar Crop Survey scheme at the end of 1943 there was rapid expansion of work at Giridih during 1944-45. The strength of staff increased from about 116 in April 1944 to more than 210 in October 1944. N. T. Mathew, Superintending Statistician, was in charge of the branch up to November 1944, and Moni Mohan Mukherjee, offg. Statistician, during the remaining part of the year.

Project and research work : The bulk of the work in the Bihar Crop Survey scheme was done at Giridih. Analysis of material collected in the Bengal Labour Enquiry during 1940-42, Sriniketan Rural Survey 1937 and data relating to blood pressure proceeded systematically throughout the year. Laboratory workers at Giridih conducted two independent local enquiries relating to family fertility and cost of living. A family budget enquiry for workers' families is in progress.

Training section : Two training courses for computers were organized at Giridih and two examinations for Computer's Certificates were held by the Institute at Giridih during the year under review.

Women's section : Work in this section proceeded satisfactorily. In the beginning most of the workers preferred part-time work, but gradually many of them became full-time workers.

Workers' Club : A social and recreational club was organized by the workers, and was running successfully for about a year. A manuscript magazine called 'Mahua' is conducted by the Club ; and an Art Exhibition was held under its auspices in June 1944. The annual sports of the Club were held in January 1945 ; and there were a number of other functions and activities.

Medical welfare scheme : In December 1944 the Honorary Secretary started a medical welfare scheme with a monthly contribution of four annas from each worker. Arrangements were made with Dr. Nirmal Kumar Das, M.B. to attend daily at the Institute and to examine each worker thoroughly and keep systematic records of such examination. Consultations at the Institute are entirely free ; visits to worker's residenec to attend the worker or any member of his family are charged for at the nominal rate of -/8/-annas per visit. The scheme has proved successful in every way and is highly appreciated by the workers.

Tiffin Club : The Tiffin Club which was started in February 1945 supplies daily tiffin to about 180 persons at a small cost. It is managed entirely by the workers of the Laboratory, and receives a subsidy from the Institute.

Puja Camp at Giridih : As in previous years the Honorary Secretary stayed at Giridih during October 1944 with a large contingent of Calcutta workers. A number of lectures and symposia were arranged at that time. The Honorary Secretary and R. C. Bose spoke on problems relating to sample surveys, and C. R. Rao on Markoff's theorem on linear estimation for several days. The Honorary Secretary met all workers on two occasions to discuss ways and means for general improvement of work at Giridih. He also entertained all workers on 23 October 1944 on the occasion of the birth-day of Mrs. Mahalanobis.

Mr. S. N. Roy, I.C.S., Financial Adviser to the Department of Civil Supplies, Bengal, and a member of the Council and Finance Committee of the Institute, visited the Giridih Laboratory during October 1944 and inspected the work with interest. Mr. Chandmull Rajgharia, Vice-President of the Institute, also visited the Laboratory in the same month.

Accommodation and construction work : With the expansion of work, lack of accommodation became an acute problem. At the end of the year the Institute was paying Rs. 705 per month as house rent. The available floor space was, however, quite inadequate ; and extensions to buildings already in the occupation of the Institute were undertaken and a good part of the work was completed during the year.

Land Acquisition : The Government of Bihar had agreed to acquire certain pieces of land adjoining the premises in the occupation of the Institute at Giridih ; notification was issued accordingly in May 1944 and enquiries by the Land Acquisition Collector were also held in June and July. No further action was, however, apparently taken by the Bihar Government for a long time, and finally without any consultation with the Institute the notification itself was cancelled, and we were informed accordingly in March 1945. We had purchased a large quantity of building materials including cement and iron, permits for which were obtained through the Government of Bihar, for the construction of buildings on lands proposed to be acquired, and a great deal of loss, due to deterioration of cement etc., is likely to be caused by the termination of the and acquisition proceedings.

Owing to increase of work in connexion with various new projects of the Government of India and also due to difficulties in securing necessary accommodation in Calcutta in the near future it will be necessary to maintain a large staff at Giridih for some considerable time. Attempts are, therefore, being made to secure additional accommodation at Giridih

SUMMER VACATION CAMP AT KALIMPONG

As in other years, a camp laboratory was set up at Kalimpong for about 2 months during the summer vacation of 1944 and was attended among others by Rajchandra Bose, Samarendranath Roy, Jitendra Mohan Sen Gupta, C. Radhakrishna Rao, Hari Kinkar Nandi, S. Janardan Poti, Dhirendranath Sarkar, and Rathindra Sen Gupta ; P. C. Mahalanobis was also there for some time. A small computing section was organized with men from Calcutta and local recruits, and considerable progress was made both in theoretical and applied topics.

STATISTICAL LIBRARY

The main stock of the Library continued to be maintained at Giridih in accordance with the general evacuation policy, and books requisitioned from Calcutta and Baranagore were sent on loan for specified periods.

Catalogue : The classification of books and the preparation of the classified catalogue with author and title index made good progress and is expected to be completed in the course of the next few months.

New books : The Library obtained import licences from Controller of Imports, Government of India, and appointed foreign agents in New York and London for importing books directly from America and U.K. Difficulties in procuring foreign books were in this way partially overcome. The total number of volumes added to the Library during the year was 858 of which 21 were received as gifts ; and the total number at the end of the year was about 25,000 volumes. Although exchange of publications with *Sankhya* has been seriously affected by the war, the Library received 174 periodicals and 22 annuals.

Issue of books : The use of the Library at Calcutta by the workers of the Laboratory, the Calcutta University and Presidency College increased very much during the year. Since December 1944 a provisional scheme was adopted to extend the Reference Library service from 8 A.M. to 8 P.M. on all week days and from 1 P.M. to 5 P.M. on Sundays. The number of volumes issued during the year was : day issues 5756, overnight 2719, loan 1435, with a total of 9910 (as against 3915 in 1943-44 and 2231 in 1942-43). Giridih and Barnagore sections of the library were used much less, the total number of volumes issued being 1127 and 254 respectively.

Circulating Library : The workers' circulating library (of general and non-technical literature in English, Bengali and Hindi) had been started at Giridih in 1942 and was extended to Calcutta and Baranagore n April 1944. In this scheme each worker paid Re 1/- as annual subscription and the Institute made a contribution of twice the amount raised as subscription from workers. The Institute also made an additional grant for the purchase of books and magazines the total number of which at the end of the year was 1891. Books were divided roughly in equal proportions at the three places and arrangements were made to interchange the books at regular intervals. This service was very popular and the total number of issues were 2624, 3029 and 2132 at Giridib, Calcutta and Baranagore respectively.

VISITORS

As usual a large number of distinguished visotors came to the Institute. His Excellency the Rt. Hon'ble Mr. Casey, C.H., D.S.O., M.C., Governor of Bengal, visited the Statistical Laboratory at Calcutta on 13 July 1944. Among other visitors mention may be made of Sir C. D. Deshmukh (Governor, Reserve Bank of India), Mr. S. Dutt, I.C.S., (Secretary, Agriculture Departent, Bengal), Mr. L. K. Elmhirst, Sir Hugh Hood (Chief Adviser to the Governor of Madras), Hon'ble Khan Bahadur S. Muazzamuddin Hossain (Minister of Agriculture, Bengal), Col. F. T. Jones, Wing Commanders Kendrew, Vermon and Barkla (R. A. F.), Major Clarke (A. E. C.) Seth Kasturbhai Lalbhai (Ahmedabad), Mr. O. M. Martin, I. C. S. (Commissioner, Postwar Reconstruction, Bengal), Mr. G. M. McKelvie (Consulting Enginneer to the Government of India), Sir Cyril Norwood, Mr. S. H. Y. Oulsnam (Jt. Secy. to the Government of India, Department of Education, Health and Lands), Dr. K. C. K. E., Raja, Sir Archibald Rowlands (Finance Member, Government of India).

Mr. M.K. Sen Gupta (Financial Adviser, Communications, Government of India), Dr. M.R. Siddiqi (Osmania University), Mr.C.O. Tattersall (Scientific Adviser, M.G.O.), Squadron Leader J.W. Vermon, Mr.R.L. Walker (Finance Secretary, Bengal) and Mr. M. W. M. Yeatts, I.C.S. (Chairman, Population Data Committee).

PROFESSOR R. A. FISHER'S VISIT

We had the great pleasure of having with us for a short time this year Prof. R. A. Fisher, F.R.S., Balfour Professor in the University of Cambridge. Arrangements for his visit were made jointly by the Institute and the British Council to which our best thanks are due for its active co-operation in this matter. He arrived in Calcutta on 24 January and stayed here till 24 February 1945. During this period he delivered a course of six lectures on Genetics in the Institute and conducted a large number of seminar meetings on statistical research which were attended by workers of the Institute and other scientific organizations in Calcutta.

Prof. Fisher's 56th birthday was celebrated by the workers of the Institute at 'Amrapali', Baranagore on 24 February 1945 and the function was attended by a large number of friends.

He visited the Cinchona Plantation at Mungpoo, and the *char* and forest areas in the Sundarbans in Bengal, and on 24 February left for Delhi, where he halted for a fortnight and left for United Kingdom by air on 11 March. In both Calcutta and Delhi he had a large number of consultations and discussions with officials and representatives of various departments of Government.

Prof. Fisher had visited India in 1938 at the invitation of the Institute. Like the first visit the second visit also was of great help to the Institute and its workers as it gave them an opportunity of having his opinion and stimulating criticisms on schemes of research. He also helped us by giving his advice about administrative aspects of statistical work.

Local Branches of the Institute

Up to the time of going to the press reports of the following branches have been received and short summaries of their activities are given below :

Poona (Deccan) Branch : Prof. V. G. Kale continued to be the Chairman of the Branch and Mr. N. V. Sovani, the local Secretary. Prof. N. V. Kanitkar delivered an interesting lecture on "the statistical lay-out of field experiments under dry farm conditions in the Bombay Deccan" on 17 January 1945. Prof. Gadgil and Mr. Sovani were engaged on the report of the socio-economic survey of Poona city during the year under review. The number of life-members during the year continued to be twelve.

Mysore Branch : Early in June 1944 the services of Prof. K. B. Madhava, Secretary of the Mysore Branch, were placed by the Mysore Government at the disposal of the Government of India who deputed him to work in the Statistical Laboratory, Calcutta. In consequence of his transfer and other changes at Mysore the activities of the Mysore Branch of the Indian Statistical Institute remained suspended.

U. P. (Lucknow) Branch : During the session 1944-45 Mr. N. C. Mehta was the president and Prof. Radhakamal Mukherjee the Secretary of the branch. Analysis of the data collected in connection with the standard of living and social conditions of the middle class in Lucknow was taken up during the year and has been nearly completed.

Delhi Branch : There were two life membrs, twelve ordinary members and forty local members. No proper meeting could be held this year. Persons who had provisionally agreed to speak had been so busy with their official work that meetings announced to be held had to be cancelled at the eleventh hour. Sir T. E. Gregory was elected as the President for the year 1945-46 and Mr. G. R. Seth, as Secretary.

Miscellaneous

Honours course in Statistics in the Presidency College, Calcutta : On the initiative of Principal A. K. Chanda and with the active collaboration of the Institute an honours degree course in Statistics was opened, for the first time in India, in the Presidency College, Calcutta, in July 1944. The inaugural address was given by P. C. Mahalanobis on the role of statistics in human affairs in general. K. B. Madhava worked as the Head of the Department in an honorary capacity ; Birendra Nath Ghose and Purnendu Kumar Bose as professor and lecturer respectively and Sunith Krishna Gupta as laboratory assistant. Out of a large number of applications 12 students were selected for admission.

Post-Graduate Department of Statistics : Our collaboration with the Post-Graduate Department of Statistics continued as before for the fourth year. P. C. Mahalanobis was the Honorary Head of the Department throughout the year ; R. C. Bose and S. N. Roy worked as lecturers, Hari Kinkar Nandi as whole-time assistant lecturer, C. Radhakrishna Rao, Nirmal Chakravorty as part-time lecturers, and Manindra

Nath Ghose and K. C. Cheriyan, as part-time assistant lecturers in the department. K. B. Madhava, Birendra Nath Ghose, Purnendu Kumar Bose and Mrs. C. Bose worked as part-time lecturers in an honorary capacity.

Eleven students appeared for the M.A. and M.Sc. examinations in Statistics held in August 1944, out of whom eight were successful. Most of the successful candidates joined the Laboratory in various capacities.

Training in Actuarial Statistics : The Institute of Actuaries, London, having announced that the examination of that Institute which had been suspended during the war would be resumed, war condition permitting, in 1945, there were inquiries whether the Indian Statistical Institute could undertake to give lectures for actuarial students. Advantage was taken of Prof. K. B. Madhava's presence in Calcutta to arrange a course of 50 lectures between August and December 1944 under the auspices of the Institute. There was considerable enthusiasm to begin with, but when the course terminated in December 1944 there were only 5 or 6 students who had attended regularly.

Bengal Administration Committee : P. C. Mahalanobis gave evidence before the Bengal Administration Committee on statistical matters. Sir Archibald Rowlands, Chairman of the Committee, and two members, Sir H. Wood, I.C.S. and Mr. R. L. Walker, I.C.S. spent one whole afternoon in the Statistical Laboratory examining the crop survey scheme and other projects.

At the request of the Administration Committee, Bengal, a memorandum prepared by K.B. Madhava and P. C. Mahalanobis was submitted by the Indian Statistical Institute on certain statisical and actuarial matters relating to recruitment to secretarial services of the Government of Bengal, giving estimates of the numbers to be annually recruited to each grade of service, current salary, costs and ultimate pension charges for cadres of different strength subjected to varying decrementary ratios.

Bengal Statistical Bureau : A Statistical Bureau, attached to the Finance Department, was started by the Government of Bengal from 15 January 1945, with P. C. Mahalanobis as part-time Statistical Adviser. No whole-time staff was appointed, but the work was done by the part-time services of the Institute staff, for which a grant of Rs. 325 for clerical work and Rs. 300 for technical help had been sanctioned by Government. The present functions of the Bureau are primarily advisory in character, e.g. the giving of technical advice regarding the collection of primary statistics, and the classification, tabulation and analysis of statistical material, and their publication.

Appreciation from abroad : The Weldon Memorial Prize was awarded by the University of Oxford to P. C. Mahalanobis in 1944, and he was elected a Fellow of the Royal Society of London in March 1945.

Audited Accounts

The accounts of the Institute were audited by Messrs P. C. Nandi & Co., Chartered Accountants. The audited accounts were considered by the Finance Committee on 25 April and the Council on 26 April, and were adopted at the Annual General Meeting held on 26 April 1945.

The Director's Contribution Fund (consisting of donations received from Prof. and Mrs. P. C. Mahalanobis and contributions received from various sources in lieu of the honorary services of P. C. Mahalanobis in connexion with project work) was separated as a special fund last year. During the financial year under review Rs. 15,000 was received from the Government of Bengal as a contribution for the honorary services of P. C. Mahalanobis in the Bengal crop survey scheme, Rs. 15,000 from the Government of Bihar for the Bihar crop survey scheme, and Rs. 1,750 from the Population Data Committee (Government of India) for the statistical work of the Committee, for which we offer our thanks to the donors. The total amount in this fund at the end of the year was Rs. 71,715-11-9.

Indian Saisical Institute : Receips and Payments Accoun for he year ending 31st March, 1945

RECEIPTS

	Rs. A. P.	Rs. A. P.
To Opening Balance	..	84,158 14 10
,, Membership fees	..	3,806 0 0
,, Interest on investments	..	3,215 10 0
,, Training fees and other receipts		
Examination fees	1,535 0 0	
Training fee	1,438 0 0	
Miscellaneous receipts	247 14 0	
Miscellaneous receipts	247 14 0	3,220 14 0
,, Block Grant : Government of India		
Department of Education, Health & Lands		53,292 0 0
Grant for Sankhya	..	2,000 0 0
Population Data Committee	..	13,250 0 0
War Transport Department	..	5,000 0 0
Controller of Food (Famine Enquiry Commission	..	2,000 0 0
To Block Grant : Government of Bengal		
Bengal Crop Survey (1944-45)	..	7,99,906 2 0
River Research Institute	..	5,000 0 0
Economic Survey of Distressed Villages	..	25,000 0 0
Controller of Rationing	..	1,559 0 0
Statistical Bureau	..	515 10 0
To Block Grant : Government of Bihar		
Bihar Crop Survey	..	2,10,000 0 0
To contribution in lieu of Prof. P. C. Mahalanobis's honorary services		
Bengal Government	15,000 0 0	
Population Data Committee	1,750 0 0	
Bihar Government	15,000 0 0	31,750 0 0
Provident fund subscription	..	8,208 15 0
Contribution to leave salary and Provident Fund	..	48,603 3 3
Income Tax realised from staff	..	284 15 0
Circulation Library	..	470 0 0
Rentals realised	..	14,692 0 0
Library deposit	..	96 0 0
		13,13,029 4 1

PAYMENTS

	Rs. A. P.	Rs. A. P.
By Salary including research staff	3,12,452 8 9	
,, Contribution to leave salary & Provident Fund	48,603 3 3	
,, Dearness Allowance	87,926 11 0	
,, Salary & Allowance (Bengal Crop Survey Field)	3,13,011 0 0	7,61,993 7 0
,, Establishment & travelling	..	35,498 7 0
,, Books & journals	..	8,064 11 2
,, Laboratory equipments	..	13,866 14 3
,, Stationery & store purchases	..	43,965 13 0
,, Postage, printing & contingency	..	21,848 13 10
,, Rentals	..	19,524 8 0
,, Other expenses (hire charges of machines, furnitures etc.)	..	34,868 5 0
,, Evacuation and emergency	..	3,200 0 9
,, Special activities at Giridih	..	886 7 0
,, Construction, maintenance & repairs	..	26,886 5 0
,, Housing Scheme at Giridih and Baranagore	..	16,832 0 0
,, Government of India grant for Sankhya	..	2,000 0 0
,, Local Branches	..	27 4 0
,, Indian Statistical Conference	..	1,207 3 0
,, Maps	..	1,236 11 0
,, Auditors Fee	..	1,500 0 0
,, Repayment of old advances to Prof. P. C. Mahalanobis		5,000 0 0
,, Examination fees	..	1,451 0 0
,, Bengal Crop Survey : Other Field expenses (1943-44)		567 13 9
,, Bengal Crop Survey : Other Field expenses (1944-45)		1,37,447 14 0
,, Bihar Crop Survey Field (1943-44)	..	11,833 10 9
,, Survey of Distressed Villages (Field)	..	19,500 0 0
,, Income Tax for Staff (1943 44)	..	1,061 4 9
,, Deposit A/C	..	2,804 6 6
,, Anthropometric Survey	..	1,000 0 0
,, Suspense a/c	..	13,119 7 3
,, Prof. R. A. Fisher's visit to India	..	3,068 6 9
,, Leave Salary & Provident Fund etc.	..	2,317 9 0
,, Circulating Library	..	3,138 7 0
Investments in Govt. Paper		
Directors Contribution Fund	71,715 11 9	
Provident Fund Subscriptions etc.	28,609 12 1	1,00,325 7 10
To Closing Balance :		
Cash	3,116 5 0	
Bank	13,870 10 5	16,986 15 5
		13,13,029 4 1

Examined and Found Correct.
P. C. NANDI & Co., *Auditors.*
Registered Accountants
Chartered Accountants.

8 HASTINGS STREET CALCUTTA,
The 25th April, 1945

AUDITED ACCOUNTS FOR 1944-45

Indian Statistical Institute : Earmarked & Special Enquiries Fund, Account for the year ending 31st March, 1945

Bengal Crop Survey

RECEIPTS	Rs.	A.	P.	PAYMENTS	Rs.	A.	P.
To Opening Balance ..	26,081	5	9	Statistical Branch (1944-45)			
" Block grant (1944-45) ..	7,99,906	2	0	By Salary & allowance ..	2,42,669	14	0
	8,25,987	7	9	By Other expenses	99,120	2	0
					3,41,790	0	0
				Field Branch			
				By Other expenses (1943-44) ..	567	13	9
				" Salary & allowance (1944-45)	3,13,011	0	0
				" Other expenses	1,23,531	4	0
				" Suspense a/c	13,916	10	0
					4,51,026	11	9
				" Closing balance	33,170	12	0
					8,25,987	7	9

Burdwan-Hoogly-Howrah Flushing Scheme

RECEIPTS	Rs.	A.	P.	PAYMENTS	Rs.	A.	P.
To Opening Balance ..	6,511	4	0	By Salary & allowance ..	4,092	14	0
				" Other expenses	1,670	2	0
					5,763	0	0
				" Closing Balance	748	4	0
					6,511	4	0

U. P. Wheat & Sugarcane Scheme

RECEIPTS	Rs.	A.	P.	PAYMENTS	Rs.	A.	P.
To Opening Balance ..	5,218	5	3	By Salary & allowance ..	2,128	7	0
				" Other expenses	868	9	0
					2,997	0	0
				" Closing balance	2,221	5	3
					5,218	5	3

Bihar Crop Survey

RECEIPTS	Rs.	A.	P.	PAYMENTS	Rs.	A.	P.
To Opening Balance (1943-44)	11,057	10	9	Statistical Branch			
" Block grant (1944-45) ..	2,10,000	0	0	By Salary & allowance (1944-45)	1,52,818	6	0
	2,21,057	10	9	Other expenses	62,418	10	0
" Advance from Indian Statistical Institute ..	6,012	15	6		2,15,237	0	0
	2,27,070	10	3	FieldBranch, other expenses (1943-44)	11,833	10	3
					2,27,070	10	3

Survey of Distressed Villages (Bengal)

RECEIPTS	Rs.	A.	P.	PAYMENTS	Rs.	A.	P.
				Statistical Branch			
To Block grant	25,000	0	0	By Salary & allowance ..	13,913	9	0
" Advance from Indian Statistical Institute ..	14,091	0	0	" Other expenses	5,677	7	0
					19,591	0	0
				Field Branch, other expenses ..	19,500	0	0
	39,091	0	0		39,091	0	0

Population Data Committee (Government of India)

RECEIPTS	Rs.	A.	P.	PAYMENTS	Rs.	A.	P.
To Block grant	13,250	0	0	By Salary & allowance ..	12,185	0	0
" Advance from Indian Statistical Institute ..	6,249	0	0	" Other expenses ..	7,314	0	0
	19,499	0	0		19,499	0	0

Road Development Enquiry (War Transport Department, Government of India)

RECEIPTS	Rs.	A.	P.	PAYMENTS	Rs.	A.	P.
To Block grant ..	5,000	0	0	By Salary & allowance ..	5,907	6	0
" Advance from Indian Statistical Institute ..	988	0	0	" Other expenses	80	10	0
	5,988	0	0		5,988	0	0

Examined and Found Correct,
P. C. NANDI & Co., *Auditors,*
Registered Accountants, Chartered Accountants.

HASTINGS STREET CALCUTTA,
The 25th April, 1945.

APPENDIX 1. LIST OF PAPERS PUBLISHED AND REPORTS SUBMITTED IN 1944-45

1. BOSE, (Mrs.) CHAMELI. The D^2 -statistic in some varietal experiment on the rice crop (*Proc. Ind. Sc. Congress*, 1945).

2. ——— Sampling error in the method of double sampling. (*Proc. Ind. Sc. Congress*, 1945).

3. BOSE, PURNENDU. Statistical study of rainfall in Calcutta : 1900-1930. (*Proc. Ind. Sc. Congress*. 1945).

4. ——— Determination of an unbiassed critical region of type A in the case of F distribution. (*Proc. Ind. Sc. Congress*, 1945).

5. BOSE, R. C. The symmetrical factorial designs of the class S^5 with S^2 plots per block. (*Proc. Ind. Sc. Congress*, 1945).

6. ——— A note on Youden's Squares. (*Proc. Ind. Sc. Congress*. 1945).

7. BOSE, R. C. & CHOWLA, S. Construction of Affine Difference Sets. (*Proc. Ind. Sc. Congress*, 1945).

8. ——— Theorems in the additive theory of numbers. (*Proc. Ind. Sc. Congress*, 1945).

9. ——— A new series of balanced incomplete block designs. (*Proc. Ind. Sc. Congress*, 1945).

10. CHAKRAVARTY, NIRMAL. On the income elasticity of demand when the incomes are distributed in the Pearsonian Type III distribution. (*Proc. Ind. Sc. Congress*, 1945).

11. ——— The present position of Indian sugar industry in relation to consumers : 1927-28 to 1939-40. (*Proc. Ind. Sc. Congress*, 1945).

12. GHOSE, BIRENDRANATH. Efficiency of rectangular plots of different shapes and sizes in field experiments or sample surveys. (*Proc. Ind. Sc. Congress*, 1945).

13. ——— The distribution of distances in a configuration of many rectangles with parallel edge. (*Proc. Ind. Sc. Congress*, 1945).

14. GHOSE, MANINDRA NATH. On the order of approximation involved in the Laplace's central limit theorem in probability. (*Proc. Ind. Sc. Congress*, 1945).

15. ——— On the problem of similar regions. (*Proc. Ind. Sc. Congress*, 1945).

16. MAHALANOBIS, P. C. On Large Scale Sample Surveys. (*Phil. Trans., Roy. Soc., London, Series B*, No. 584, 231, 329-451).

17. ——— Report on the Bihar Crop Survey : Rabi Season 1943-44. (Submitted to Government of Bihar in November 1944).

18. ——— Note on per capita consumption of food stuff in Bengal, 1936-1942. (Submitted to the Famine Inquiry Commission in December 1944).

19. ——— Note on the mortality in Bengal in 1943. (Submitted to the Famine Inquiry Commission in February 1945).

20. ——— Bengal Crop Survey 1944-45 : Progressive reports were submitted to the Government of Bengal on jute and *aus* paddy crops on 31 August, 16 September and 14 October 1944 ; and on *aman* paddy crop on 19 December 1944 and 16 and 24 January and 1, 8, 14 and 23 February 1945.

21. ——— Bihar Crop Survey 1944-45 : Progressive reports on *bhadoi* (monsoon) crops were submitted on 8 and 20 October and 4 November 1944 ; on *aghani* (winter) crops on 11 November, 1, 8 and 25 December 1944 ; 2 and 22 January, 2 and 9 February, and 21 March 1945 ; and on *rabi* (spring) crops on 19 and 23 March 1945.

22. NANDI, HARIKINKAR. On the average power of test criteria. (*Proc. Ind. Sc. Congress*, 1945).

23. ——— Note on the enumeration of all possible non-isomorphic solutions of balanced incomplete block designs. (*Proc. Ind. Sc. Congress*, 1945).

APPENDIX 1. (*Contd.*)

24. Rao, C. Radhakrishna. On finite geometries and theorems of the number theory derivable from them. (*Proc. Ind. Sc. Congress*, 1945).

25. ——— Some new cyclic solutions of incomplete block designs. (*Proc. Ind. Sc. Congress*, 1945).

26. ——— Generalisation of Markoff's Theorems and tests of linear hypothesis. (*Proc. Ind. Sc. Congress*, 1945).

27. Rao, C. Radhakrishna. On linear estimation theory and testing of hypothesis. (*Current Science*, 13(6) June, 1944).

28. ——— Familial correlations or the multivariate generalisation of the intra-class correlation. (*Current Science*, March 1945).

29. ——— Extension of the difference theorems of Singer and Bose (*Science and Culture*, 10(1), July, 1944).

30. ——— On the linear set up leading to intra and inter-block information. (*Science and Cuture* - 10(6), December, 1944).

31. Roy Samarendranath. The individual sampling distribution of p-statistics on the non null hypothesis when the null hypothesis concerns equality of mean values for *k* p-variate normal. population. (*Proc. Ind Sc Congress*, 1945)

APPENDIX 2. OFFICERS ON DEPUTATION FOR TRAINING OR SCIENTIFIC CONSULTATION : 1944-45

1. Q. M. Hossain M.Sc. (*Lecturer in Physics & Statistics, Dacca University, Dacca*), 13 to 16 April, 1944.
2. O. S. Nayar, M.Sc., (*Lecturer in Mathematics, Jorhat College, Assam*) 2 May to 24 June 1944.
3. P. R. Roy, M.Sc., (*Statistician, Department of Agriculture, Bengal*) 5 to 13 May 1944.
4. J. R. Pal, B.Ag. (*Assistant Rice Reasearch Officer, Bankura*) 13 to 20 May, 19 to 22 July, 29 July, 4 to 10 August 1944, and 2 to 6 March 1945.
5. Das Moni Roy, M.A. (*Professor of Mathematics, Meerut College, U.P.*) 12 July to 11 December 1944.
6. M. Y. Kazi, M.A.(*Department of Economics Muslim University, Aligarh*) 22 July to 10 October, 1944.
7. K.P. Pillai, M.Sc.(*Physiological Assistant, Ravenshaw College, Cuttack*) 2 May 1944, 5 to 12 March 1945.
8. Subrata Sen. M.Sc. (*Dacca Agricultural Farm, Dacca*) 9 to 31 October, 1944.
9. Raghav Prosad Mishra, B.Sc. (*Sugarcane Research Station, Pusa*) 21 October 1944 to 5 January 1945.
10. Amode Ranjan Sen, M.Sc. (*Senior Statistical Assistant, Department of Agriculture, U.P. Lucknow*) 2 January to 28 February, 1945.
11. Suprakash Sen, M.Sc. (*Imperial Veterinary Research Institute, Mukteswar-Kumaon, U.P.*) 2 to 22 January 1945.

APPENDIX 3. STUDENTS IN THE TRAINING SECTION. 1944-45.

(1) Y. D. Borwanker, M.A. (*Allahabad*); (2) Robi Narayan Das, B.Sc., (*Patna*) ; (3) Girija Sankar Prosad, M.A. (*Patna*) ; (4) S. Visuvalingam, B.A. (London) (*Ceylon*), ; (5) P. Sundara Rama Avadhany, B.A. (*Madras*); (6) Bindhyeshwari Pd. Singh, M.A. (*Patna*) ; (7) Dayal Prasad, M.A.(Math.) M.A. (Econ.) (*Agra*) ; (8) Khursheed Alam M.A. (*Patna*) ; (9) Salil Bindu Ghose, M.A. (*Lucknow*); all from August 1944. (10) Prosad Kumar Banerjee, M.Sc. (*Calcutta*) and (11) Anadi Ranjan Roy, M.A. (Math.) & M.A. (Statistics) (*Calcutta*) in October and November 1944.

INDIAN STATISTICAL INSTITUTE
Proceedings of the Council dated 26 April 1945

A meeting of the Council of the Indian Statistical Institute was held at 5-0 P.M. on Thursday the 26th April 1945, in the Statistical Laboratory, Presidency College, Calcutta.

In the absence of the President, Sir C. V. Raman (Vice-President) took the chair.

Present : Mrs. Chameli Bose, Prof. K. N. Chakravarti, Mr. N. C. Chakravarti, Prof. H. C. Ghose, Mr. B. Das Gupta, Prof. K. B. Madhava, Mr. S. N. Ray, I.C.S., Mr. Samarendranath Roy and Prof. P. C. Mahalanobis (Honorary Secretary).

1. Sir C. V. Raman proposed from the chair that the Council notes with gratification the election of Prof. P. C. Mahalanobis as a Fellow of the Royal Society.

2. Proceedings of the meeting of the Council dated 23 February 1945 were read and confirmed.

3. *Election of members* : Narendra Nath Sarkar, M.A. (Chief Accountant, Corporation of Calcutta) was elected life member and the following persons were elected ordinary members of the Institute :

Purnendu Kumar Bose, M.Sc. (Statistical Laboratory, Calcutta), Birendranath Ghosh, M.Sc. (Statistical Laboratory, Calcutta), Dhirendra Mohan Ganguli, M.Sc. (Dy. Superintendent, Bengal Crop Survey, Calcutta), Mahabirprasad Roy, M.Sc., (Calcutta Corporation), Palghat Ramakrishna Subramanyam, M.A. (Stock Share and Finance Broker, Bombay).

4. *Bengal Crop Survey* : The Honorary Secretary reported that the Government of Bengal (Agriculture Dept. Memo No. 3425 dated 4 April 1945) had sanctioned a grant of Rs. 6,50,000 plus dearness allowance of Rs. 1,50,000 for the continuance of the Bengal Crop Survey during 1945-46. Mr. Nihar Chandra Chakravarti, B.C.S., Asst. Secretary, Agriculture Department, would continue to be in charge of the Field Branch. The preparatory work had been completed and the field survey had already started.

Resolved that the Honorary Secretary be authorised to take all necessary action in this matter.

5. *Appointment of Committees* : Resolved that the following Committees be re-constituted as follows, for one year, with powers to co-opt :

Journal Committee : Mr. Rajchandra Bose, Prof. Satyendranath Bose (Dacca), Dr. C. Chandra Sekar, Mr. Haricharan Ghose, Prof. K. B. Madhava, Prof. P. C. Mahalanobis (Editor), The Hon'ble Mr. Justice T. J. Y. Roxburgh and Mr. Samarendranath Roy (Secretary).

Finance Committee : The President, The Treasurer, The Secretary, Prof. Khagendranath Chakravarti (Convener) Mr. Sudhir Kumar Banerjee, Mr. Nihar Chandra Chakravarti, Mr. Haricharan Ghosh, Mr. Benoy Bhusan Das Gupta, Prof. K. B. Madhava, Mr. Satyendranath Ray, I.C.S. and Mr. Samarendranath Roy.

6. *Examinations Committee* : Resolved that Prof. Satyendranath Bose, Prof. K. B. Madhava and Prof. P. C. Mahalanobis be elected representatives of the Council on the Examinations Committee.

7. *Recommendations from the Finance Committee* : Considered the following extracts from the proceedings of the Finance Committee dated 25 April 1945 :

"*Land Acquisition at Baranagore* : The Hon. Secretary placed before the Council a letter from the Land Acquisition Officer, and from the owner of the land, a letter offering to sell the land by private negotiation. The Hon. Secretary explained the present position, and stated that he was trying to settle the matter by private negotiation. He suggested that expenditure for purchasing the land may be met by borrowing, with his consent, from the Director's Contribution Fund.

Resolved that the Secretary be authorised to take necessary action on above lines and keep the Finance Committee informed as to progress of negotiation.

It was also recommended to the Council that necessary money be borrowed from the Director's Contribution Fund with the consent of Prof. Mahalanobis for the purpose of acquiring the land and that arrangements be made for payment of reasonable interest and for the ultimate repayment of the borrowed sums to the Fund."

Resolved that the proposal made by the Finance Committee be approved.

8. *Budget Estimate for* 1945-46 : Considered the draft budget estimate for the year 1945-46 as prepared by the Finance Committee.

Resolved that the budget as modified by discussion in the Council be passed.

Budget Estimate for 1945-46

	Rs.	A.	P.
Opening Balance (Cash & Bank)	16,986	15	5
Grant from the Govt. of India ..	51,000	0	0
Grant from the Govt. of India for Sankhya ..	2,000	0	0
Membership fees	1,000	0	0
Govt. of Bengal : Crop Survey	8,00,000	0	0
Govt. of India : War Transport	75,000	0	0
Govt. of India : Population Data Committee	20,000	0	0
Govt. of Bihar · Crop Survey ..	42,533	0	0
Contribution in lieu of Prof. Mahalanobis's honorary services	25,000	0	0
Miscellaneous Receipts ..	1,480	0	7
Rentals	14,000	0	0
	10,49,000	0	0

	Rs.	A.	P.
Salary including research staff..	3,10,000	0	0
Establishment & travelling ..	25,000	0	0
Books and journals ..	10,000	0	0
Laboratory equipments	7,500	0	0
Stationery & stores purchase ..	30,000	0	0
Postage, printing & contingency	15,000	0	0
Rentals	20,000	0	0
Other expenses (hire charges of machines, furniture etc.)	30,000	0	0
A.R.P. evacuation & emergency	3,000	0	0
Special activities at Giridih ..	2,000	0	0
Construction, maintenance & repairs ..	5,000	0	0
Housing scheme ..	10,000	0	0
Govt. of India grant for Sankhya	2,000	0	0
Contribution to Sankhya from Institute	3,000	0	0
Director's Contribution Fund ..	25,000	0	0
Indian Statistical Conference ..	1,000	0	0
Bengal Crop Survey (Field) ..	5,39,000	0	0
Auditor's fee	1,500	0	0
Closing Balance ..	10,000	0	0
	10,49,000	0	0

9. *Annual Report* : The draft Annual Report for the year 1944-45 was placed before the meeting by the Hon. Secretary.

Resolved that the draft Thirteenth Annual Report be forwarded to the Annual General Meeting.

10. *Audited Accounts* : The Hon. Secretary placed before the Council the Audited Report for 1944-45 which had been considered by the Finance Committee on 25 April 1945 and also statements relating to project and other funds.

Resolved that the audited accounts and statements relating to funds be forwarded to the Annual General Meeting.

11. *Institute Examinations* : To suit the convenience of examiners as well as of the examinees resolved that the Institute examinations this year be postponed till July/August 1945 and the Hon. Secretary be authorised to fix the dates.

Resolved also that Prof. K. B. Madhava, Prof. P. C. Mahalanobis (Secretary), Mr. R. C. Bose, and Mr. S. N. Roy be appointed members of the Board of Examiners for conducting the examination for the Computer's Certificate proposed to be held sometime in July/August and that the above four examiners together with Dr. P. V. Sukhatme, Mr. K. R. Nair and Dr. C. Chandra Sekar be appointed members of the Board of Examiners for the examination for the Statistician's Diploma to be held in July/August 1945.

12. *Membership fees of Local Branches* : In modification of first part of Bye Law (VI) of Bye-laws relating to Local Branches (adopted by the Council on 3 August 1932) the membership fee of Rs. 20 for members resident within the area covered by a Local Branch be allocated as follows :—

(a) Rs. 13 for the Journal (a copy of which will be supplied free to each member of the Institute).

(b) Rs. 3-8-0 for the Local Branch Fund.

(c) Rs. 3-8-0 for the Institute Fund.

13. *Local Branch at Ahmedabad* : Considered a letter from Mr. G. D. Mathur dated Ahmedabad, 20 April 1945 proposing that a Branch of the Institute be started at Ahmedabad with the following persons as office-bearers :—

(1) *President* : Mr. Kasturbhai Lalbhai; (2) *Treasurer* : Mr. C. V. Mavlankar; (3) *Secretary* : Mr. G. D. Mathur; (4) *Joint Secretaries* : Profs. Y. D. Keskar and J. B. Sandil.

Resolved that sanction be given to the opening of a local Branch at Ahmedabad in accordance with the constitution of the Indian Statistical Institute and the list of office-bearers as above be approved.

14. *Over-draft arrangements* : The Honorary Secretary stated that owing to non-receipt of sanctioned project grants it was necessary to meet current expenses by increasing the limit of the over-draft of Rs. 25,000 arranged with the Imperial Bank of India.

Resolved that the limit of the over-draft arrangement with the Imperial Bank of India be increased to Rs. 50,000 for a period of 3 months.

Also resolved that the Finance Committee be requested to examine the questions of the working balance.

15. *Land Acquisition arrangements at Giridih* : The Hon. Secretary stated that the Government of Bihar had decided, without consultation with the Institute, to terminate the Bihar Sample Survey of Crops at the end of the work in the *rabi* season of 1944-45 and also the land acquisition procedure for acquiring land at Giridih for the Institute. He also stated that Prof. R. A. Fisher, F.R.S. was of opinion that the discontinuance of the sample survey of crops in Bihar would be a retrograde step as this was likely to create a serious gap in crop statistics; the Hon. Secretary was also of the same opinion. The Hon. Secretary further stated that the Institute was in need of land at Giridih not merely for the Bihar crop survey work but also for other projects and its general purposes, and that, with the approval and support of the Government of Bihar, iron, cement and other building materials for putting up structures on land proposed to be acquired had been purchased at Giridih. A good deal of this material was liable to be wasted and thus involve the Institute in heavy pecuniary loss.

Resolved that the Hon. Secretary be authorised to make suitable representations to the Government of Bihar in this matter.

C. V. Raman
Chairman.

P. C. Mahalanobis
Honorary Secretary.

INDIAN STATISTICAL INSTITUTE

Proceedings of the Annual General Meeting dated 26 April 1945

The Annual General Meeting of the Indian Statistical Institute was held at 5-45 p.m. on Thursday, the 26th April 1945, in the Statistical Laboratory, Presidency College, Calcutta.

Sir C. V. Raman, D.Sc., F.R.S., N.L., (*Vice President*), was in the chair.

The following members were present :—

Dr. B. N. Bhaduri, M.Sc., D.I.C., Ph.D. (Lond.)	*Asst. Physicist, New Research & Control Laboratory Jamshedpur.*
Mr. H. G. Bhattacharjee, M.A., B.Com.	*Statistical Section, D. G. M. P.'s Office, Dept. of Supply, Calcutta.*
Mrs. C. Bose, B.Sc. (Lond.)	*Statistical Laboratory, Calcutta.*
Prof. K. N. Chakravarti, M.Sc.	*Presidency College, Calcutta.*
Mr. N. C. Chakravarti, M.A., B.C.S.	*Addl. Asst. Secretary to the Govt. of Bengal, Agriculture Dept.*
Mr. B. Das Gupta, M.A.	*Secretary, Indian Central Jute Committee.*
Prof. H. C. Ghosh, M.A., P.R.S.	*Calcutta University.*
Prof. K. B. Madhava, M.A., A.I.A.	*Mysore University.*
Mr. S. N. Ray, I.C.S.,	*Financial Adviser, Dept. of Civil Supplies, Bengal.*
Mr. Samarendranath Roy, M.Sc.	*Calcutta University.*
Prof. P. C. Mahalanobis, F. R. S.	*Honorary Secretary.*

1. Sir C. V. Raman proposed from the chair that the Indian Statistical Institute notes with gratification the election of Prof. P. C. Mahalanobis as a Fellow of the Royal Society.

2. *Annual Report and Audited Accounts* : The Honorary Secretary placed before the meeting the Annual Report and Audited Accounts for 1944-45 as approved by the Council and mentioned that the Accounts had been audited by Messrs P. C. Nandi & Co., Chartered Accountants.

Resolved that the Annual Report and Audited Accounts for 1944-45 be adopted.

(*Carried unanimously*).

Proposed by Prof. K. B. Madhava
Seconded by Mrs. Chameli Bose

3. *Election of Office-Bearers and Members of the Council* : The Honorary Secretary reported that in accordance with the Bye-law 3 for election of office-bearers and members of the Council, nominations for 1945-46 approved by the Council of the Institute had been circulated together with the notice of the Annual General Meeting on the 9th April, 1945. As the number of nominations was in no case in excess of the number of vacancies, no voting was necessary and no voting paper was therefore issued.

The election of the following office-bearers and members of the Council as circulated was then announced from the chair.

President

Sir C. D. Deshmukh, C.I.E.	*Governor, Reserve Bank of India.*

Hon. Vice-Presidents (having filled the office of President)

Sir Badridas Goenka, Kt,. C.I.E.,	*Calcutta.*
Mr. N. R. Sarker	*Calcutta.*

Vice-Presidents

Dr. P. N. Banerjea, M.A., D.Sc., M.L.A. (Central).	*Calcutta.*
Sir A. H. Ghuznavi, Kt., M.L.A. (Central).	*Messrs A. H. Ghuznavi & Co., Calcutta.*
Sir T. E. Gregory, D.Sc.,	*Economic Adviser to the Government of India, Delhi.*
Dr. S. P. Mookerjee, M.A., B.L., M.L.A., D.Litt., Bar-at-law.	*Calcutta.*
Sir C. W. B. Normand, Kt. C.I.E., M.A.,D.Sc.	*Special Officer, The Observatory, Delhi.*
Mr. Chandmull Rajgharia	*Mica Merchant, Giridih.*
Sir Shri Ram, Kt.	*Delhi Cloth and General Mills, Delhi.*
Sir C. V. Raman, D.Sc., F.R.S., N.L.,	*Indian Institute of Science, Bangalore.*
The Hon'ble Mr. Justice T. J. Y. Roxburgh B.A., Bar-at-law, C.I.E., I.C.S.	*Judge, Calcutta High Court.*
Mr. H. S. Suhrawardy	*Calcutta.*

Treasurer

Dr. Satya Churn Law, M.A., B.L., Ph.D., F.Z.S., M.B.O.U.	*Calcutta.*

Secretary

Prof. P. C. Mahalanobis, F.R.S., ...	*Statistical Laboratory, Calcutta.*

Joint Secretaries

Prof. K. N. Chakravarti, M.Sc.	*Presidency College, Calcutta.*
Kumar Bimal Chandra Sinha, M.A., ..	*Calcutta.*

Members of the Council

Dr. S. K. Banerjee, D.Sc.	*Director General of Observatories, Delhi.*
Mr. Sudhir Kumar Banerjee	*Statistical Laboratory, Calcutta.*
Mrs. C. Bose, B.Sc., (London)	*Statistical Laboratory, Calcutta.*
Mr. Raj Chandra Bose, M.A., F.N.I., ..	*Statistics Department, Calcutta University.*
Prof. S. N. Bose, M.Sc.	*University of Dacca.*
Mr. P. H. Carpenter, O.B.E., F.I.C., F.C.S. ..	*Director, Indian Tea Association, Tocklai, Assam.*
Mr. Nihar Chandra Chakravarti, M.A, B.C.S...	*Addl. Asst. Secretary, Government of Bengal. Department of Agriculture.*
Dr. C. Chandra Sekar, M.Sc. Ph.D. ..	*All India Institute of Hygiene and Public Health, Calcutta.*
Prof. K. P. Chattopadhyay, M.Sc. (Cantab), ..	*University of Calcutta.*
Mr. B. Das Gupta, M.A., ...	*Secretary, Indian Central Jute Committee.*
Prof. H. C. Ghose, M.A., P.R.S., ...	*University of Calcutta.*
Prof. F. W. Levi, Dr. Phil. Nat.,.. ..	*University of Calcutta.*
Prof. John Maclean, M.A.	*Wilson College, Bombay.*
Mr. A. C. Mukherji, M.A., Dip. de Statisticien, F.S.S.	*Director of Statistics, Baroda State, Baroda.*
Mr. K. R. Nair, M.A., M.Sc., ...	*Food Department, Government of India, Delhi.*
Dr. K. C. K. E. Raja, L.R.C.P. & S., D.P.H., D.T.M. & H., ...	*Deputy Public Health Commissioner, Delhi.*
Dr. V. K. R. V. Rao, M.A., Ph.D. ..	*Director of Statistics, Food Department, Delhi.*
Mr. S. N. Ray, B.A. (Cantab), I.C.S. ..	*Financial Adviser, Dept. of Civil Supplies, Bengal.*
Mr. Samarendranath Roy, M.Sc., F.N.I. ..	*University of Calcutta.*
Dr. P. V. Sukhatme, Ph.D., D.Sc. ..	*Statistician, Imperial Council of Agricultural Research, Delhi.*

4. *Appointment of Auditors* : Resolved that Messrs P. C. Nandi & Co., Chartered Accountants be appointed auditors for the year 1945-46 on an honorarium of Rs. 1500.

Proposed by Mr. N. C. Chakravarti
Seconded by Prof. H. C. Ghosh

(*Carried unanimously*).

5. *Appointment of Committees* : Resolved that as recommended by the Council the following Committees be reconstituted as follows with powers to co-opt.

Journal Committee (1 year) : Mr. Rajchandra Bose, Prof. Satyendra Nath Bose (Dacca), Dr. C. Chandra Sekar, Prof. Haricharan Ghosh, Prof. K. B. Madhava, Prof. P. C. Mahalanobis (*Editor*). The Hon'ble Mr. Justice T. J. Y. Roxburgh and Mr. Samarendranath Roy (*Secretary*).

Finance Committee (1 year) : The President, the Treasurer, the Secretary, Mr. Sudhir Kumar Banerjee, Prof. K. N. Chakravarti (Convener), Mr. N. C. Chakravarti, Mr. B. Das Gupta, Prof. H. C. Ghosh, Prof. K. B. Madhava, Mr. S. N. Ray, I.C.S., and Mr. Samarendranath Roy.

Proposed by Dr. B. N. Bhaduri
Seconded by Mrs. Chameli Bose

(*Carried unanimously*).

6. Vote of thanks to the retiring Office-bearers :

Resolved that the Council places on record its appreciation of services rendered by retiring office-bearers and members of the Council.

Proposed by Prof. P. C. Mahalanobis
Seconded by Prof. K. B. Madhava

(*Carried unanimously*)

C. V. Raman
Chairman.

P. C. Mahalanobis
Honorary Secretary.

SANKHYĀ
THE INDIAN JOURNAL OF STATISTICS

Edited By P. C. MAHALANOBIS

VOL. 7 PART 3 1946

ON THE LINEAR COMBINATION OF OBSERVATIONS AND THE GENERAL THEORY OF LEAST SQUARES

BY C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta.

INTRODUCTION

The general theory of linear combinations of observations as estimates of parametric functions has been treated by the author (Rao : 1943, 1945a, 1945b) under very general conditions. A unified method of approach to the problem of testing of linear hypotheses, involved in a variety of cases, has also been put forward and the necessary distributions worked out in the case when the variances and covariances are known and observations, whose expectations are linear functions of certain parameters, form a multivariate normal system. This problem which admits a simple solution leading to the use of published tables of χ^2 alone brings in fresh complications when the variances and covariances are not known. But by suitable 'Studentisation' it has been possible to show (Rao : 1945c) that many problems in testing of linear hypotheses, for some of which solutions were not available, can be solved with the help of two distributions viz Fisher's t and z alone.

The main aim of the present paper is to bring out the generality of the method of least squares of which all tests of linear hypotheses come out as special cases. Various aspects of this problem have been considered by Markoff (1912), Student (1908), Fisher (1922, 1925) Neyman and Pearson (1930, 1931) and David and Neyman (1939). A recent paper which gives a useful application of the general ideas is due to Kolodziejczyk (1935) who following the theories of Neyman and Pearson (1930,1931,1933) gave a general treatment of the problem but the conclusions arrived at are same as those of the above authors. *The important restrictions in the problem treated above are certain inequality relations connecting the number of parameters, the number of observations and the rank of the matrix of the compounding coefficients of parameters. These restrictions have now been withdrawn and it has been possible to show that by a certain rule of procedure the necessary statistics can be obtained from minimised sum of squares.* This problem for some particular cases of linear hypothesis has been treated by Bose (1944) from a different point of view.

The problem treated here is the theory of estimation and testing of hypothesis concerned with the linear functions of certain parameters and regression coefficients the compounding coefficients in the latter case being functions of concomitant variates. The theory of least squares may now be made secure on solid foundations and a unique principle, 'without which the solution of new problems has, as it were, to be picked out, afresh, by intution' leading to a 'generality in the customary methods' is made available.

As an application of this method the appropriate linear hypothesis and the analysis of variance and covariance in biological experiments have been considered and a general theory of statistical regression is given. Some examples are worked out to illustrate the theory.

In the general treatment of the problem some results are stated without proof as they can be easily constructed from the methods given by Rao (1945a, 1945b)

2. THE GENERAL PROBLEM WITH CONCOMITANT VARIATES.

Let $Y = (y_1\ y_2 \dots y_n)$ be the vector of n stochastic variates $y_1, y_2, \dots y_n$ and X be the matrix, with n rows and k columns, of k concomitant variates, the k observations $(i=1,2,\dots,k)$ in the j-th row corresponding to the j-th stochastic variate y_j. Let $\Theta = (\theta_1, \theta_2 \dots \theta_n)$ be the vector of expectation of Y for the given set of observations on the concomitant variates. This is expressed as $E[Y : X] = \Theta$ or simply $E(Y) = \Theta$ when there is no ambiguity. It is given that

$$\Theta = FG' = TA' + RX' \qquad \dots (2{\cdot}1)$$

where F and G are partitional matrices defined by $F = (T \mid R)$ and $G = (A \mid X)$, $T = (\tau_1\ \tau_2 \dots \tau_m)$ and $R = (\rho_1\ \rho_2 \dots \rho_k)$ are the row matrices of m unknown parameters τ's and k unknown parameters ρ's and A is a known matrix with n rows and m columns.

The assumption made above is one of linear regression on the concomitant variates and this can be done, without loss of generality, for if the regression is a general polynomial or any other linear combination of functions of directly observable concomitant variates (such as squares, cubes etc. in the case of polynomial regression or logorithms, exponential etc.) each of these functions may be regarded as a separate concomitant variate. The general problem treated here is the case where, for a given set of observations on concomitant variates the expectation of stochastic variates are linear functions of m unknown parameters τ's and k unknown parameters ρ's, the compounding coefficients for the ρ parameters being single valued functions of concomitant variates.

If $y_1, y_2, \dots, y_n$ do not have any functional relationship we can get a positive definite matrix Λ of rank n as $E[Y-\Theta)'(Y-\Theta) : X]$ where E stands for the mathematical expectation, in which case Λ is referred to as the dispersion matrix of the stochastic vector Y for the observed concomitant matrix X. It is assumed that the elements of Λ are finite and are known apart from a constant multiplier. *We are not assuming any equality on inequality relations among n, k, m and the ranks of A and X.* The set of equations $E(Y) = FG'$ are known as observational equations.

There are two types of problems to be answered in the theory of linear estimation connected with concomitant variates.

(i) Given a linear parametric fonction such as LT' where $L = (l_1\ l_2 \dots l_m)$ is a row matrix, the problem is to find a vector B such that

(a) $E(BY') = LT'$ independently of T and R $\qquad \dots (2.2)$

(b) $V(BY')$ is minimum. $\qquad \dots (2.3)$

(ii) Given a linear parametric function such as MR' where $M = (m_1 m_2 \dots m_k)$ is a row matrix the problem is to find a vector C such that

(a) $E(CY') = MR'$ independently of T and R $\qquad \dots (2.4)$

(b) $V(CY')$ is minimum. $\qquad \dots (2.5)$

In the case (a) is satisfied the corresponding parametric functions are said to be estimable and if (b) is also satisfied the parametric functions are said to have the best unbiassed estimates given by the corresponding linear functions of the variates.

The problems mentioned above are only particular cases of the general problem considered by Rao (1945a) but requires a special discussion in view of the importance of the role of concomitant variates which will be dealt with in a later section.

The definition of unbiassedness involved in (2·2) gives that if

$$LT' = E\,(BY') = BAT' + BXR' \tag{2.6}$$

then

$$BA = L \text{ and } BX = 0 \tag{2.7}$$

Also if B exists such that $BA = L$ and $BX = 0$ then $LT' = E(BY')$ which gives the following theorem.

Theorem 2a. *The necessary and sufficient condition that LT' is estimable is that there exists a vector B such that $BA = L$ and $BX = 0$.*

From the set of B's satisfying the condition (2.7) we have to choose B such that $V(BY') = BAB'$ is the least. Introducing the vectors $2D = (d_1\, d_2\, ..d_m)$ and $2F = (f_1\, f_2\, ..f_k)$ of Lagrangian multipliers and minimising the expression.

$$BAB' - 2D(A'B' - L') - 2FX'B' \tag{2·8}$$

We get

$$BA - DA' - FX' = 0 \tag{2·9}$$

$$BA = L, \qquad BX = 0 \tag{2·10}$$

Multiplying both sides of (2.9) by A^{-1}, the inverse of A, we get the vector B leading to the best estimate as

$$B = DA'A^{-1} + FX'A^{-1} \tag{2.11}$$

which with the help of (2.10) shows that the vectors D and F satisfy

$$DA'A^{-1}A + FX'A^{-1}A = L \tag{2·12}$$

$$DA'A^{-1}X + FX'A^{-1}X = 0 \tag{2·13}$$

the best estimate and the minimum variance being

$$BY' = DA'A^{-1}Y' + FX'A^{-1}Y' \tag{2.14}$$

$$DA'A^{-1}AD' + FX'A^{-1}AD' = LD' \tag{2·15}$$

Theorem 2b. *The necessary and sufficient condition that the parametric function LT' is estimable is that there exist vectors D and F satisfying the equations*

$$(D\,|\,F)(A'\,|\,X')A^{-1}\;A = L$$

$$(D\,|\,F)(A'\,|\,X')A^{-1}\;X = 0$$

and the best unbiassed estimate of LT', if estimable, and its variance are given by $(D\,|\,F)(A'\,|\,X')A^{-1}Y'$ and LD' respectively.

The necessary and sufficient condition given in theorem 2b is only another form of that given in theorem 2a and is more useful as it is directly connected with the best estimate. Similarly we have the following theorems regarding the estimability and the best estimate of the parametric functions.

Theorem 2c. *The necessary and sufficient condition that the parametric function MR' is estimable is that there exists a vector C such that $CA = O$ and $CX = M$.*

Theorem 2d. *The necessary and sufficient condition that MR' is estimable is that there exist vectors G and H satisfying the equations*

$$(G\,|\,H)\;(A'\,|\,X')A^{-1}A = O$$

$$(G\,|\,H)\;(A'\,|\,X')A^{-1}X = M$$

and the best estimate of MR' *if estimable and its variance are given by* $(G|H)(A'|X')\Lambda^{-1}Y'$ *and* MH' *respectively.*

The following theorem on uniqueness of the results can be proved as in (Rao : 1945a.)

Theorem 2e. *The best unbiassed estimates are unique independently of whatever may be the vectors F and D of theorem 2b and G and H of theorem 2d satisfying the corresponding equations and the variances of the estimates thus arrived at are the least.*

3. EXTENSION OF MARKOFF'S RESULTS

In this section a simple and a practical procedure for the problem of estimation of linear parametric functions, without laying much emphasis on the algebra given in the previous section has been developed. From the variates of the stochastic vector Y we construct the variates $Y\Lambda^{-1}(A|X)$ by a linear transformation so that

$$E[Y\Lambda^{-1}(A|X)]=(T|R)\,(A|X)'\Lambda^{-1}(A|X) \qquad \text{.. (3.1)}$$

leading to the equations

$$E(U)=(T|R)H'_1,\; E(W)=(T|R)H'_2 \qquad \text{.. (3·2)}$$

where

$$U=Y\Lambda^{-1}A,\; W=Y\Lambda^{-1}X \qquad \text{.. (3.3)}$$

and

$$(A|X)'\,\Lambda^{-1}(A|X)=(H_1|H_2)' \qquad \text{.. (3.4)}$$

From theorem 2b we get that if there exist vectors F and D such that

$$LT'\equiv(F|D)\,(H_1|H_2)\;(T|R)' \qquad \text{.. (3.5)}$$

then the best estimate of LT' is given by $(F|D)\;(U'|W')$. If $(\bar{T}|\bar{R})=(t_1\,t_2\ldots t_m|r_1\,r_2\ldots r_k)$ be a solution of the equations $(U|W)=(T|R)\,(H_1|H_2)'$ then, because of (3·5) we have

$$L\bar{T}'=(F|D)\,(H_1|H_2)\,(\bar{T}|\bar{R})'=(F|D)\;(U'|W') \qquad \text{.. (3·6)}$$

which shows that the best estimate of LT' if estimable, can be obtained from $L\bar{T}'$ where $\bar{T}=(t_1t_2\ldots t_m)$ as defined above. This is unique independently of whatever may be the solution $(\bar{T}|\bar{R})$ satisfying the above equations. Following the proof given in (1945a) we can show that the result of substituting $(\bar{T}|\bar{R})$ in LT' leads to the best estimate if and only if

(i) $L\bar{T}'$ is homogeneous in y's

(ii) $E(L\bar{T}')=LT'$

are satisfied. This shows that the parametric function LT' need not be tested for estimability before applying Markoff's principle of substitution. If it is not estimable either one or both of the conditions are violated in which case its estimate ceases to exist ; similar results can be derived in the case of the parametric function MR'. Hence we get the following theorem.

Theorem 3a. *The best estimate of the parametric function* LT' *(or* MR'*) exists and is given by* $L\bar{T}'$ *(or* $M\bar{R}'$*) where* $\bar{T}$ *(or* $\bar{R}$*) is the vector giving a set of solutions of the equations defined in* (3·1) $Y\Lambda^{-1}(A|X)=(T|R)(A'\Lambda^{-1}A+X'\Lambda^{-1}X)$ *if (i)* $L\bar{T}'$ *(or* $M\bar{R}'$*) is homogeneous in the stochastic variates and (ii)* $E(L\bar{T}')=LT'$ *(or* $E(M\bar{R}')=MR'$*).*

Since $(\bar{T}|\bar{R})$ is any solution of the equations given above we get the following result.

Theorem 3b. *The vector* $(\bar{T}|\bar{R})$ *mentioned in theorem 3a can be obtained after adding any consistent and convinient set of equations in* τ*'s and* ρ*'s to the equations given in theorem (3a) viz.* $Y\Lambda^{-1}(A|X)=(T|R)\,(A'\Lambda^{-1}A+X'\Lambda^{-1}X)$ *and solving them.*

The equations $(U|W)=(T|R)\,(H_1|H_2)'$ are called normal equations and are readily obtained by equating to zeroes the partial derivatives of

$$\sum_i\sum_j \lambda^{ij}\,(y_i-\theta_i)\,(y_j-\theta_j) \qquad \text{.. (3.7)}$$

when λ^{ij} are the elements of Λ^{-1}, with respect to the parameters τ's and ρ's respectively.

The analysis is complete with the help of the following theorem which has been proved in Rao (1945a: 1945b).

Theorem 3c. *The set of equations* $Y\Lambda^{-1}(A\,|\,X)=(T\,|\,R)\,(A'\Lambda^{-1}A+X'\Lambda^{-1}X)$ *are always solvable for* τ*'s and* ρ*'s.*

4. REGRESSION PARAMETERS AND THE STRUCTURE OF CONCOMITANT VARIATES.

The parameters $\rho_1, \rho_2, \ldots, \rho_k$, whose compounding coefficients are functions of concomitant variates, are known as regression parameters. In this Section some properties of regression parameters, simpler methods of estimation and the structure of concomitant variates are discussed.

From the normal equations $(U\,|\,V)=(T\,|\,R)\,(H\,|\,H_2)'$ we can, by the method of sweep out, eliminate the set of parameters $\tau_1, \tau_2, \ldots \tau_m$ or $\rho_1, \rho_2, \ldots \rho_k$ in which case the resulting equations may be represented as

$$P=TS' \quad \text{or} \quad Z=RN' \qquad \ldots \quad (4\cdot1)$$

It has been shown (Rao : 1945a) that the dispersion matrix of $(U\,|\,V)$ is $(H_1\,|\,H_2)$ which is the matrix of the normal equations and this intrinsic property of the normal equations is preserved even when some of the parameters are eliminated by the method of sweep out. Hence we get the following theorem.

Theorem 4a. *The dispersion matrices of P and Z, of the equations* $P=TS'$ *and* $Z=RN'$ *obtained from the normal equations* $(U\,|\,V)=(T\,|\,R)(H_1\,|\,H_2)'$ *by the method of sweep out, are the matrices of resulting equations S and N respectively.*

The theorem 4a gives us that if the best estimate of LT' (or MR') is BP' (or CZ') then

$$V\,(BP')=BSB'=LB' \qquad \ldots \quad (4.2)$$

$$V\,(CZ')=CNC'=MC' \qquad \ldots \quad (4.3)$$

and if L_1T' and L_2T' are estimated by B_1P' and B_2P' then

$$\text{Cov}\,(B_1P'\,.\,B_2\,P')=L_1B'_2=L_2B'_1 \qquad \ldots \quad (4.4)$$

with similar results for the covariance of the best estimates of M_1R' and M_2R'.

We could by any other device eliminate the τ's and get the equations in ρ's but the above intrinsic properties are not usually preserved. But in this case we can construct another system of parameters which are linear functions of ρ's and adjust such that the matrix of the equations in the new variables is the dispersion matrix of the corresponding quantities which occur on one side of the equations. By certain devices we can preserve these intrinsic properties and these are particularily important as the further calculation of standard errors are simplified.

The normal equations from which the best estimates are to be obtained are

$$Y\Lambda^{-1}A = TA'\,\Lambda^{-1}A+RX'\,\Lambda^{-1}A \qquad \ldots \quad (4.5)$$

$$Y\Lambda^{-1}X = TA'\,\Lambda^{-1}X+RX'\Lambda^{-1}X \qquad \ldots \quad (4.6)$$

If the regression parameters are zero then the normal equations for the τ parameters corresponding to the observational equations $E(Y)=TA'$ are

$$Y\Lambda^{-1}A=TA'\Lambda^{-1}A \qquad \ldots \quad (4.7)$$

Since the parametric functions TA' are all estimable there exists a matrix F with n rows and m columns such that

$$TA'\equiv TA'\Lambda^{-1}AF' \qquad \ldots \quad (4.8)$$

in which case the best estimates of the parametric functions TΛ′ are given by $Y\Lambda^{-1}AF'$. Since (4.8) holds identically

$$A' = A'\Lambda^{-1}AF' \qquad .. \quad (4.9)$$

Premultiplication of both sides of (4.9) by F and taking the transpose we get

$$FA' = FA'\Lambda^{-1}AF' \qquad .. \quad (4.10)$$

$$AF' = FA'\Lambda^{-1}AF' \qquad .. \quad (4.11)$$

which shows

$$AF' = FA' \qquad .. \quad (4.12)$$

Post multiplication of both sides of (4.5) by F′ we get

$$Y\Lambda^{-1}AF' = TA'\Lambda^{-1}AF' + RX'\Lambda^{-1}AF' \qquad .. \quad (4.13)$$

$$YN' = TA'N' + RX'N' \qquad .. \quad (4.14)$$

where

$$N' = \Lambda^{-1}AF'$$

Post multiplying both sides of (4.14) with $\Lambda^{-1}X$ and subtracting from (4.6) we get the equations giving the regression parameters as

$$Y(I-N')\Lambda^{-1}X = RX'(I-N')\Lambda^{-1}X \qquad .. \quad (4.15)$$

which is the same as

$$Y(I-N')\Lambda^{-1}(I-N)X = RX'(I-N')\Lambda^{-1}(I-N)X \qquad .. \quad (4.16)$$

since

$$\begin{aligned}(I-N')\Lambda^{-1}N &= \Lambda^{-1}N - \Lambda^{-1}AF'\Lambda^{-1}FA'\Lambda^{-1}\\ &= \Lambda^{-1}N - \Lambda^{-1}FA'\Lambda^{-1} = \Lambda^{-1}N - \Lambda^{-1}N = 0 \qquad .. \quad (4.17)\end{aligned}$$

by the use of (4.10) and (4.12). Also the dispersion matrix of $Y(I-N')\Lambda^{-1}X$ is

$$X'\Lambda^{-1}(I-N)\Lambda(I-N')\Lambda^{-1}X = X'(I-N')\Lambda^{-1}X \qquad .. \quad (4.18)$$

for

$$\begin{aligned}X'\Lambda^{-1}N\Lambda(I-N')\Lambda^{-1}X &= X'\Lambda^{-1}FA'(I-N')\Lambda^{-1}X\\ &= X'\Lambda^{-1}(FA'-AF') = 0 \qquad .. \quad (4.19)\end{aligned}$$

which shows that by the method adopted above the variance-covariance property is preserved in the equations (4.15) for the estimation of regression parameters. The set of equations in the form expressed in (4.16) shows that these are normal equations derivable from the set of observational equations $E(\eta) = R\xi'$ with the dispersion matrix of η as Λ where

$$\eta = Y(I-N') \text{ and } \xi = (I-N)X \qquad .. \quad (4.20)$$

which may be called the *residual* stochastic vector and the concomitant matrix respectively.

Let $X_1, X_2, \ldots, X_k$ represent the vectors of the 1st, 2nd, ... and k-th concomitant variates, the i-th element of each corresponding to y_i the i-th stochastic variate. We can construct the observational equations

$$E(Y) = T^{(o)}A' \text{ and } E(X_i) = T^{(i)}A' \qquad .. \quad (4.21)$$

$$(i = 1, 2, \ldots k)$$

where $T^{(i)} = (\tau_1^{(i)}\ \tau_2^{(i)} \ldots \tau_m^{(i)})$ are the parameters corresponding to the i-th concomitant vector and and $T^{(o)} = (\tau_1^{(o)}\ \tau_2^{(o)} \ldots \tau_m^{(o)})$ for the stochastic vector and obtain the best unbiassed estimates of the parametric functions $T^{(o)}A'$ and $T^{(i)}A'$. There are n variables in each case and n best estimates of their corresponding expectations. If we denote the best estimates by placing a bar over the vector of the parameters we get the residual vectors

$$\eta = Y - \bar{T}^{(o)}A' \text{ and } \xi_i = X_i - \bar{T}^{(i)}A' \qquad .. \quad (4.22)$$

mentioned in (4·18). Hence we get the theorem.

Theorem 4b. *Tke normal equations leading to the best estimates of the regression parameters are derivable from the observational equations* $E(\eta) = R\xi'$ *construsted with the help of residual vectors* η *and* ξ_i *derived in* (4·22) *and assuming the dispersion matrix of* η *to be* Λ. *The equations thus obtained for the estimation of regression parameters possess the variance-covariance properties.*

With the help of these equations $\eta\,\Lambda^{-1}\xi = R\xi'\,\Lambda^{-1}\xi$ we can solve for the p parameters and use it for the purpose of estimating linear functions of regression parameters and finding their variances and covariances. Having obtained the vector $\bar{R}$ of solutions we can substitute it in (4·5) to obtain the equations in the τ parameters.

$$Y\,\Lambda^{-1}\,A - \bar{R}X'\,\Lambda^{-1}A = TA'\,\Lambda^{-1}A \qquad .. \quad (4.23)$$

If LT′ is an estimable paramatric function then there exists a vector C such that $CA'\Lambda^{-1}AT' = LT'$ in which case the best estimate is

$$C\,A'\Lambda^{-1}Y' - CA'\,\Lambda^{-1}X\,\bar{R}' = L\bar{T}^{(\omega)\prime} - \Sigma r_i L\bar{T}^{(i)\prime} \qquad .. \quad (4.24)$$

where $\bar{T}^{(\omega)}$ and $\bar{T}^{(i)}$ are as defined above and r_i is solution for ρ_1 the ith regression parameter. Hence we get the theorem.

Theorem 4c. *The best unbiassed estimate of an estimable parametric function LT' is given by* $LT^{(\omega)\prime} - \Sigma r_i\,L\bar{T}^{(i)\prime}$ *where* $LT^{(\omega)\prime}$ *and* $L\bar{T}^{(i)\prime}$ *are the best estimates of* $LT^{(\omega)\prime}$ *and* $LT^{(i)\prime}$ *from the observational equations* $E(Y) = T^{(\omega)}A'$ *and* $E(X_i) = T^{(i)}A'$ *and* r_i *are the solutions of the equations* $\eta\Lambda^{-1}\xi = R\xi'\Lambda^{-1}\xi$ *where* η *and* ξ *are the residual vectors defined in* (4·22).

It appears as if the parametric functions connected with the stochastic variates and the concomitant variates are related with the parametric function LT′ in the manner

$$LT' = LT^{(\omega)\prime} - \Sigma\rho_i\,LT^{(i)\prime} \qquad .. \quad (4.25)$$

which may admit a physical interpretation with reference to particular problems. We can find the variances and covariances of the estimates given in (4.24) with the help of the following theorem. The expressions $L\bar{T}^{(i)\prime}$ are constants being only functions of the concomitant variates.

Theorem 4d. *The function* $L\bar{T}^{(\omega)\prime}$ *the estimate of* LT' *when the regression parameters are zero is uncorrelated with the estimates of linear functions of regression parameters.*

The proof follows from the fact that $L\bar{T}^{(\omega)\prime}$ is a linear combination of $Y\Lambda^{-1}A$ or $Y\Lambda^{-1}AF' = YN'$ defined above, and the estimate of a linear function of the regression parameters is a linear combination of $Y\Lambda^{-1}(I - N')$ in which case the relation $(I - N')\Lambda^{-1}N = 0$, proved in (4.17) showing that the elements of YN' and $Y\Lambda^{-1}(I - N')$ are uncorrelated, establishes the theorem. Hence we have

$$V[L\bar{T}^{(\omega)\prime} - \Sigma r_i\,L\bar{T}^{(i)\prime}] = V[L\bar{T}^{(\omega)\prime}] + V[\Sigma r_i\,L\bar{T}^{(i)\prime}] \qquad .. \quad (4·26)$$

for each of which the variances are known.
Similarly

$$\begin{aligned} &\mathrm{Cov}[L\bar{T}^{(\omega)\prime} - \Sigma r_i\,L\bar{T}^{(i)\prime}]\,[M\bar{T}^{(\omega)\prime} - \Sigma r_i\,M\bar{T}^{(i)\prime}] \\ &= \mathrm{Cov}[L\bar{T}^{(\omega)\prime} \,.\, M\bar{T}^{(\omega)\prime}] + \mathrm{Cov}[\Sigma r_i\,L\bar{T}^{(i)\prime} \,.\, \Sigma r_i\,M\bar{T}^{(i)\prime}] \end{aligned} \qquad .. \quad (4·27)$$

The expressions (4.25) for the estimate of a parametric function LT′ and the expressions (4.26) and (4.27) for the variances and covariances consist of two portions (i) the corresponding expressions when the regression parameters are absent and (ii) a correction due to the concomitant variates. The calculations are so arranged that the expressions obtained by assuming the regression parameters to be zero are made use of in deriving the correction factors themselves.

It is interesting to note that the above analysis is not limited to concomitant variates alone but applicable whenever we want to distinguish two sets of parameters in the theory of linear estimation. Thus we could, in the above case, reverse the roles of τ's and ρ's and consider the elements of A as observations on certain concomitant variates.

5. THE CASE OF INDEPENDENT VARIATES

A general problem that occurs in the estimation of regression and other parameters is when the stochastic variates are independent and have the same common variance. If this common variance is σ^2 then on replacing Λ by σ^2 I, in the estimating equations we find that σ^2 cancells out in the normal equations giving the result that the best estimates are independent of the value of σ^2 whether known or unkonwn. In this section the practical procedure for estimation is dealt with at length and certain short cuts are suggested.

Let the observational equations be

$$E(y_i)=a_{i1}\tau_1+\dots+a_{im}\tau_m+\rho_1 x_{1i}+\dots+\rho_k x_{ki} \qquad i=1,2\dots,n \qquad \dots \tag{5·1}$$

with certain restrictions on parameters

$$g_i=r_{i1}\tau_1+\dots+r_{im}\tau_m \qquad i=1,2,\dots,s \qquad \dots \tag{5·2}$$

In the general treatment given above we have not considered linear restrictions on the parameters but as shown in (Rao : 1945b) this does not introduce any fresh difficulty but in deriving the normal equations the quadratic form :

$$\lambda^{ij}(y_i-\theta_i)(y_j-\theta_j) \qquad \dots \tag{5·3}$$

has to be minimised subject to the linear restrictions. The expressions for the estimates, variances and covariances all follow from the general treatment given in (Rao : 1945b).

Assuming the regression parameters to be zero we obtain the best estimates of parametric functions. Let $c_{i0}+c_{i1}y_1+\dots+c_{in}y_n=\bar{y}_i$ be the best estimate of $a_{i1}\tau_1+\dots+a_{im}\tau_m$ when ρ's are zero but τ's subject to (5.2). Then the i-th residual is $(y_i-\bar{y}_i)$. Similarly the i-th residual for the j-th concomitant variate is $\xi_{ji}=x_{ji}-c_{i0}-c_{i1}x_{j1}-\dots-c_{in}x_{jn}$ where, c's are the same as those used in the case of y's. The residual observational equations are then

$$E(\eta_i)=\rho_1\xi_{1i}+\dots+\rho_k\xi_{ki} \qquad i=1,2,\dots,n \qquad \dots \tag{5·4}$$

from which we get the normal equations giving ρ's as

$$\Sigma\eta_i\xi_{ji}=\rho_1\Sigma\xi_{1i}\xi_{ji}+\dots+\rho_k\Sigma\xi_{ki}\xi_{ji} \qquad j=1,2,\dots,k \qquad \dots \tag{5·5}$$

Certain short cuts are available if solutions of the normal equations corresponding to each of the variates are available. Starting from the observational equations (5.1) with the restrictions (5.2) and assuming that the regression parameters are absent we get the normal equations whose solution we represent by $t_1^{(o)},t_2^{(o)},\dots,t_m^{(o)}$. By replacing the n observations on the stochastic variate by those of the j-th concomitant variate we can get corresponding solutions which we represent by $t_1^{(j)},t_2^{(j)},\dots,t_m^{(j)}$.

$$\Sigma\eta_i\xi_{ji}=\Sigma(y_i-a_{i1}t_1^{(o)}-\dots)(x_{ji}-a_{i1}t_1^{(j)}-\dots.)$$

$$=\Sigma y_i x_{ji}-t_1^{(j)}\Sigma a_{i1}y_1-t_2^{(j)}\Sigma a_{i2}y_i-\dots. \qquad \dots \tag{5·6}$$

$$=\Sigma y_i x_{ji}-t_1^{(o)}\Sigma a_{i1}x_{ji}-t_2^{(o)}\Sigma a_{i2}x_{ji}-\dots. \qquad \dots \tag{5·7}$$

Similarily

$$\Sigma\,\xi_{im}\xi_{jm}=\Sigma x_{im}x_{jm}-t_1^{(j)}\Sigma a_{m1}x_{im}-\dots\dots\dots \qquad \dots \tag{5·8}$$

$$=\Sigma x_{im}x_{jm}-t_1^{(i)}\Sigma a_{m1}x_{jm}-\dots\dots\dots \qquad \dots \tag{5·9}$$

With the help of (5.6) to (5.9) we can set down the quations (5.5) from which the regression parameters are solved for. If $r_1, r_2, \ldots, r_k$ are the solutions thus obtained then the best estimate of $l_1\tau_1 + \ldots + l_m\tau_m$, if estimable, is given by

$$\Sigma l_1(t_1^{(o)} - r_1 t_1^{(1)} - r_2 t_i^{(2)} - \ldots\ldots) \qquad \ldots \quad (5\cdot10)$$

The estimates $\Sigma l_1 t_i^{(o)}$ and $\Sigma\Sigma l_1 r_j t_i^{(j)}$ are uncorrelated as shown in theorem 4d, so that the variance of (5.10) is the sum of the variances of its two uncorrelated parts. The variance of each part is easily derivable from the normal equations from which the $t_1^{(o)}$ and r_j are obtained. Thus if $t_1^{(o)}$ are obtained from the equations.

$$Q_i = b_{i1} t_1^{(o)} + \ldots + b_{im} t_m^{(o)} \qquad \ldots \quad (5\cdot11)$$
$$s = 1, 2, \ldots, m$$

and (5.2) and if $\Sigma l_1 t_i^{(o)} = b_o + \Sigma b_i Q_i$ then $V(\Sigma l_1 t_i^{(o)}) = \Sigma l_1 b_1 \sigma^2$ and if the equations giving the regression parameters are

$$P_i = h_{io} + h_{i1} r_1 + \ldots + h_{ik} r_k \qquad \ldots \quad (5\cdot12)$$
$$i = 1, 2, \ldots k$$

and if $\Sigma m_i r_1 = n_o + \Sigma n_i P_i$ then $V(\Sigma m_1 r_i) = \Sigma m_1 n_i \sigma^2$ and so on for the covariances also . It is to be observed here that the solutions t's and r's represent the best estimates of the parameters τ's and ρ's only when they are estimable individually but they are helpful otherwise in giving the best estimates of the estimable parametric functions. Estimability can also be tested with the help of these solutions as indicated before. All the equations used above are always solvable (Rao : 1945a) and it is enough for our purpose to get any single solution.

6. THE GENERAL THEORY OF LEAST SQUARES

The hypothesis involved in the theory of linear estimation is the assignment of the value of a single parametric function or the values of a number of parametric functions. The method of construction of suitable statistics and their distributions when the stochastic variates form a multivariate normal system and the variances and covariances are known have been discussed fully in Rao (1945a, 1945b). If LT′ is a parametric function whose assigned value is ξ then denoting the estimate of LT′ by $L\bar{T}'$ we construct the statistic

$$v = (L\bar{T}' - \xi)/\sqrt{V(L\bar{T}')} \qquad \ldots \quad (6\cdot1)$$

and use it as a normal variate. To test the hypothesis $L_1T' = \xi_1, \ldots, L_rT' = \xi_r$ concerning r independent parametric functions we find the vector of best estimates $P = (P_1\ P_2 \ldots P_r)$ where $P_i = L_i\bar{T}' - \xi_i$ of $L_1T' - \xi_1, \ldots, L_rT' - \xi_r$ with their dispersion matrix D and construct the generalised variance statistic V defined by the root of the determinental equation

$$|P'P - VD| = 0 \qquad \ldots \quad (6\cdot2)$$

The method of deriving the generalised variance statistic is to take a linear compound $\lambda_1P_1 + \ldots + \lambda_r P_r$ of $P_1, P_2, \ldots P_r$ and maximise the statistic $\Sigma\lambda_i P_i/\sqrt{V(\Sigma\lambda_i P_i)}$. The statistic V is distributed as χ^2 with r degrees of freedom as shown in Rao (1945a, 1945b)

Lemma 1. *The generalised variance statistic V designed to test the hypothesis* $L_1T' = \xi_1, \ldots, L_rT' = \xi_r$ *concerning r independent parametric functions and derived as a root of the determinental equation obtained by the principle of maximisation is invariant under linear transformations of the hypothesis or the variates and is distributed as* χ^2 *with r degrees of freedom when the variances and covariances are known.*

When the variances and covariances are not known the tests can be performed only after studentising the above statistics. The general method of studentisation and the derivation of statistics in the case the stochastic variates are uncorrelated and have the same variance have been discussed here and the other cases are reserved for a subsequent communication. We may now state the problem as follows.

Given $Y=(y_1\, y_2 \cdot\cdot y_n)$ *the vector of n stochastic variables, such that* $E(Y)=TA'$, *where A is a known matrix with n rows and m columns and* $T=(\tau_1\, \tau_2 \ldots \tau_m)$ *is the vector of m unknown parameters subject to the restrictions* $G=TR'$ *where G is a row vector and R is a matrix with q rows and m columns, and further that* Λ *the dispersion matrix of Y is of the form* $\sigma^2 I$, *where* σ^2 *is unknown, the problem is to test the hypothesis whether the r independent parametric functions* $L_1T', \ldots, L_rT'$ *have the assigned values* $\xi_1, \ldots, \xi_r$. *Some of the parameters* τ*'s may be regression parameters in which case their coefficients are functions of concomitant variates.*

Following the procedure given above let $P=(P_1 \cdot\cdot P_r)$ be the vector of best estimates of $L_1T'-\xi_1, \cdot\cdot, L_rT'-\xi_r$. Since the stochastic variates are all uncorrelated the estimating vector P is independent of the unknown parameter σ^2. The dispersion matrix of P can in this case be written as $\sigma^2 D$ where the matrix D is completely known. The generalised variance statistic then comes out as

$$|P'P-V\sigma^2 D|=0 \qquad \cdot\cdot \qquad (6\cdot3)$$

$$V=\frac{PD^{-1}P'}{\sigma^2} \qquad \cdot\cdot \qquad (6\cdot4)$$

where D^{-1} is the inverse of D and the unknown prameter σ^2 appears only in the denominator.

If the y's form a normal system then V is distributed as χ^2 with r degrees of freedom when the null-hypothesis is correct. On the other hand if we can get the totality of linear functions of y's whose expectations are known independently of the null hypothesis then the generalised variance statistic associated with them is distributed as χ^2 with certain number of degrees of freedom. If $Q=(Q_1 \ldots Q_s)$ are s independent functions of y's whose expectations ars known to be $\zeta_1, \ldots, \zeta_s$ and if the dispersion matrix of Q's is F then the generalised variance statistic

$$V_0=\frac{(Q-\zeta)\, F^{-1}(Q-\zeta)'}{\sigma^2} \qquad \cdot\cdot \qquad (6\cdot5)$$

is distributed as χ^2 with s degrees of freedom. If V_0 is distributsd independently of V which is distributed as χ^2 with r degrees of freedom on the null hypothesis then a studentised statistic can be constructed by forming a function of V_0 and V which will be independent of σ^2 the simplest of which is the ratio V/V_0. The statistic sV/rV_0 is distributed as the variance ratio with r and s degrees of freedom respectively. The tests of significance can in this case be carried out with exactitude. We will now show that such an exact test is available on the maximum possible degrees of freedom for V_0 and the exact expressions for V and V_0 can be obtained by the simple device of least squares.

Theorem 6a. *If the ranks of A, R and* $(A'|R')$ *are s, k and p respectively then there are* $n+k-p$ *independent functions of the form* c_0+CY' *whose expectations are identically zeroes.*

Evidently $(s+k)\geqslant p$ and $d=(s+k-p)$ gives the number of independent rows of R which can be derived as linear combinations of the rows in A. If $E(c_0+CY')=0$ then $c_0+CAT'=0$ subject to the condition $G=TR'$ which shows that there exists a vector D such that

$$CA+DR=0 \text{ and } c_0=DG' \qquad \cdot\cdot \qquad (6\cdot6)$$

Since the rank of A is s, there are $n-s$ independent vectors C which satisfy (6.6) when $D=0$. When D is non-null there are d vectors satisfying (6.6). This follows from the definition of

d given above. The value of c_o for each of the C's is given by DG′ .Hence the number of independent functions whose expectations are zeres is

$$n-s+d=n+k-p \tag{6 7}$$

Theorem 6b. *If c_o+CY' is any function whose expectation is zero and b_o+BY' is the function giving the best unbiassed estimate of a parametric function LT' then $BC'=0$ or the two functions are uncorrelated.*

It has been shown elsewhere (Rao : 1945b) that the best estimates of parametric functions are obtained from the normal equations

$$YA=TA'A+SR \text{ and } G=TR' \tag{6 ·8}$$

where $S=(\sigma_1\ \sigma_2\ldots\sigma_q)$ is a vector of q pseudo-parameters. If LT′ is estimable then there exist two vectors χ_1 and χ_2 such athat

$$\chi_1 A'A+\chi_2 R=L \tag{6 ·9}$$

$$\chi_1 R'=0,\ b_o=\chi_2 G' \tag{6 ·10}$$

in which case the best estimate of LT′ is given by b_o+BY' where $B=\chi_1 A'$. If $E(c_o+CY')=0$ then from (6.6) we get $CA+DR=0$. Premultiplying by χ_1 we get

$$\chi_2 A'C'+\chi_1 R'D'=0=\chi_1 A'C'=BC' \tag{6 ·11}$$

Hence the theorem 6b follows.

Theorem 6c. *If the best estimates of n parametric functions TA' are represented by $\bar{T}A'$ then the expectation vector of the residual vector $Y-\bar{T}A'$ is null and the totality of the linear functions of y's (not necessarily homogeneous) whose expectations are zeroes are derivable by linear combinations of the elements of the residual vector alone.*

Let $Y\Gamma'+\Delta$ (where Γ' is a matrix with n rows and m columns and Δ is a vector with n elements) give the best estimates of TA′. Since $E(Y-Y\Gamma'-\Delta)=TA'-TA'=0$, we get that the expectation of the residual vector $Y(I-\Gamma')-\Delta$ is null. By theorem 6b we also get that $\Gamma(I-\Gamma')=0$. It can be easily shown that the number of estimable functions for which the estimates are not pure constants is $s-d$ where s and d are as defined in theorem 6a, in which case it follows that the rank of Γ is $(s-d)$ and because of the relation $\Gamma(1-\Gamma')=0$ it follows that the rank of $(1-\Gamma')$ is $(n-s+d)$. Further any vector D such that $D(1-\Gamma')=0$ makes $D\Delta'=0$ for otherwise the expectation of the residual vector is not zero. Thus the residual vector $Y(1-\Gamma')-\Delta$ constitutes $(n-s+d)$ independent functions whose expectations are zeroes and by theorem 6a this gives the totality of such functions.

Theorem 6d. *The generalised variance statistic V_o derived from these $(n-s+d)$ independent functions whose expectations are zeroes is given by $[Y(I-\Gamma')-\Delta][(I-\Gamma)Y'-\Delta']/\sigma^2$ and is distributed as χ^2 with $(n-s+d)$ degrees of freedom independently of the functions giving the best estimates of parametric functions.*

Instead of taking the number of independent functions we can take a linear function $B[(I-\Gamma)Y'-\Delta')]$ of the m functions in $Y(I-\Gamma')-\Delta$ with its variance $B(I-\Gamma)(1-\Gamma')B'\sigma^2$ and maximise the statistic V_o given by

$$B[(1-\Gamma)Y'-\Delta']\,[Y(1-\Gamma')-\Delta]=V_o\,B(1-\Gamma)B'\sigma^2 \tag{6 ·12}$$

Differentiating with B we get

$$B[(1-\Gamma)Y'-\Delta']\,[Y(1-\Gamma')-\Delta]=V_o\,B(1-\Gamma)\sigma^2 \tag{6 ·13}$$

Multiplying the above expressions by themselves we get

$$V_o\,\sigma^2=[Y(1-\Gamma')-\Delta]\,[(1-\Gamma)Y'-\Delta'] \tag{6 ·14}$$

This is evidently distributed as χ^2 with $(n-s+d)$ degrees of freedom as it is the generalised variance statistic constructed out of $(n-s+d)$ independent functions. This is distributed independently of the best estimates by theorem 6b.

Theorem 6e. *The expression $V_0\sigma^2$ (which is independent of σ^2) is the same as the minimum value of $(Y-TA')(AT'-Y')$ when minimised with respect to the τ's subject to the condition $G=TR'$.*

The expression $V_0\sigma^2$ in (6.14) is the sum of squares of the residuals $(Y-\bar{T}A')(A\bar{T}'-Y')$ where $\bar{T}$ is a any solution of the normal equations which are obtained by equating the partial derivatives of $(Y-TA')(AT'-Y')$ with respect to τ's subject to the condition $G=TR'$ to zeroes. Hence the theorem.

The above theorem gives an easy method of evaluating $V_0\sigma^2$. Now to test the hypothesis $L_1T'=\xi_1, L_2T'=\xi_2, \ldots, L_rT'=\xi_r$ we have to calculate the statistic $V\sigma^2=PD^{-1}P'$ given in (6.4) and use the statistic $V\sigma^2(n-s+d)/V_0\sigma^2 r$ as the variance ratio with r and $(n-s+d)$ degrees of freedom. We may now observe that if $L_1T'=\xi_1, \ldots L_rT'=\xi_r$, then r more linear functions of y's have zero expectations so that the generalised variance appropriate to totality of independent functions, $(n-s+d+r)$ in number, can be obtained with the help of theorem 6e by finding the minimum value of $(Y-TA')(AT'-Y')$ subject to the conditions $G=TR'$ and the given relations $L_1T'=\xi_1, \ldots, L_rT'=\xi_r$. Let V_1 be its value. This generalised variance can also be obtained starting from the best estimates $P_1, P_2, \ldots, P_r$ of $L_1T'=\xi_1, \ldots, L_rT'-\xi_r$ and the residual vectors $Y(I-\Gamma')-\Delta$ which are all uncorrelated with P's due to theorem 6b. The generalised variance of the latter alone is V_0. If the generalised variance due to two unecorrelated sets of functions are additive then we infer that the generalised variance due to P's alone is V_1-V_0. We now prove the following theorems.

Theorem 6*f*. *If $Q_1=(q_{11}\, q_{12} \cdots q_{1r})$ and $Q_2=(q_{21}\, q_{22} \cdots q_{2m})$ are two sets of functions such that no function of the first set is correlated with those of the other and have dispersion matrices $\sigma^2\Lambda_1$ and $\sigma^2\Lambda_2$ then V, the generalised variance of $(Q_1 | Q_2)$ is equal to the sum of V_1 the generalised variance of Q_1 and V_2 the generalised variance of Q_2.*

If L_1 and L_2 are the vectors of compounding co-efficients we get V as the solution of

$$L_1Q'_1Q_1+L_2Q'_2Q_1=VL_1\Lambda_1 \qquad \ldots (6\cdot15)$$

$$L_1Q'_1Q_2+L_2Q'_2Q_2=V L_2\Lambda_2 \qquad \ldots (6\cdot16)$$

from which by proper multiplications and addition we get

$$(V_1+V_2)(L_1Q'_1+L_2Q'_2)=V(L_1Q'_1+L_2Q'_2) \qquad \ldots (6\cdot17)$$

or

$$V_1+V_2=V \qquad \ldots (6\cdot18)$$

Theorem 6g. *The generalised variance V appropriate to the best estimates of $L_1T'-\xi_1, \ldots, L_rT'-\xi_r$ is obtained as the difference between the minimum values of $(Y-TA')(AT'-Y')$ when minimised with respect to the parameters of the vector T subject to the restriction $G=TR'$ and when minimised with the further restrictions $L_1T'-\xi_1=0, \ldots L_rT'-\xi_r=0$.*

The proof of this theorem follows from the results of theorem 6f and the discussion above.

This leads us to the following procedure in the theory of least squares. The observational equations are $E(Y)=TA'$ with the restrictions $G=TR'$. The hypothesis to be tested may be put as $H=TF'$. The hypothesis can be tested when and only when the system of equations $H=TF'$ are not inconsistent with $G=TR'$ and the matrix F' can be obtained from $(A'|R')$ by linear combination of its row or columns. If the minimum values of $(Y--TA')(AT'-Y')$ subject to $G=TR'$ and $G=TR'$, $H=TF'$ are denoted by $V_0\sigma^2$ and $V_1\sigma^2$

then V_0 and V_1 are distributed as χ^2 with $(n-s+d_1)$ and $(n-s+d_1+d_2)$ degrees of freedom respectively where d_1 is the number of independent vectors in R derivable from A and d_2 is the additional number independent vectors in F derivable from A and s is the rank A. If the ranks of A, R, $(A'|R')$, $(R'|F')$ and $(A'|R'|F')$ are s, k, p, r and t then $d_1=s+k-p$ and $d_1+d_2=s+r-t$. In any particular problem these degrees of freedom can be obtained from other simpler considerations. To test the hypothesis $H=TF'$ the statistic $\dfrac{V_1-V_0}{d_2}\cdot\dfrac{n-s+d_1}{V_0}$ is used as the variance ratio with d_2 and $(n-s+d_1)$. degrees of freedom.

The minimum values can be uniqualy obtained from

$$(Y-\bar{T}A')\,(A\bar{T}'-Y') \qquad \text{.. } (6\cdot18)$$

where $\bar{T}$ is any solution from the equations obtained by equating the partial derivatives of $(Y-TA')(AT'-Y')$ subject to the restraining conditions to zero. We thus arrive at the general theory of least squares without making any restrictions on the number of variables or the parameters or the rank of the matrix of observational equations. This covers the case of independent stochastic variates whose expectations are linear functions of regression and certain other parameters, the coefficients of regression parameters being functions of concomitant variates.

Since the generalised variance statistic V_0 appropriate to s independent linear functions whose expectations are known independently of the hypothesis is distributed as χ^2 with s degrees of freedom it follows that $E(V_0)=s$. This gives us another method of calculating the number of such independent linear functions or the degrees of freedom associated with V_0. We need calculate only the expectation value of the minimised sum of squares $\sigma^2V_0=(Y-\bar{T}A')(Y'-A\bar{T}')$ where $\bar{T}$ is the vector of solutions from normal equations. If the parametirc functions TA' are estimated by $QF'+\Delta$ where $Q=YA$, F is a matrix with n rows and m columns and Δ is a row matrix with n elements then

$$\begin{aligned}\sigma^2V_0&=(Y-QF'-\Delta)\,(Y-FQ'-\Delta)\\ E(\sigma^2V_0)&=E(Y-\Theta)\,(Y'-\Theta')-E[QF'-E(QF')]\,[FQ'-E(FQ')]\\ &=\sigma^2(n-\text{trace of}AF')\end{aligned} \qquad \text{.. } (6\cdot19)$$

where the trace of AF' is the sum of the diagonal elements of the matrix AF'. This follows from the fact that the variance of $CQ'+D$ the estimate of LT', a single parametric function is given by $LC'\sigma^2$. Hence we get the following theorem.

Theorem 6*h*. *If the parametric functions TA' are estimated by $QF'+\Delta$ then the degrees of freedom of V_0 the minimised sum of squares $(Y-TA')(Y'-AT')$ subject to assigned restrictions is given by $(n-$trace of $AF')$*

7. COVARIANCE TECHNIQUE IN FIELD AND BIOLOGICAL EXPERIMENTS

As an application of the previous analysis we may consider the analysis of data from biological and field experiments. Certain functions of observations from these experiments have their expectations as linear functions of unknown parameters. The nature of the linear hypothesis involved in such problems and the construction of suitable functions as observational equations amenable to reduction by the method of least squares have been discussed elsewhere (Rao : 1944). The mathematical treatment of the question is carried out after a translation of the particular problem into mathematical language with special reference to field experiments.

The problem is to test whether one variety is superior to another in a certain specified region. We cannot utilise the whole of the specified region nor is it desirable to choose certain number of plots at random for one variety and a certain number for the other. Evidently two varieties cannot be tried one after the other on identical plots without vitiating the comparisons. So we choose sets of plots adjacent to one another which are called blocks. Let us consider a block of k plots and imagine that the yields on these k plots with respect to a hypothetical variety are $\alpha_1, \alpha_2, \ldots, \alpha_k$. We define the quantities

$$\beta=(\alpha_1+\alpha_2+\ldots+\alpha_k)/k,\ \eta_i=\alpha_i-\beta,\ \Sigma\eta_i=0 \qquad \ldots (7\cdot1)$$

If any given variety is tried on the i-th plot then the observed yield y may be supposed to have the composition

$$y=\alpha_i+\tau+\varepsilon=\beta+\tau+\eta_i+\varepsilon \qquad \ldots (7\ 2)$$

where ε is a technical error due to experimentation and is such that its average over repeated experimentation under similar conditions is zero. The parameter τ may be called the *effect* of the variety. If $x_1, x_2, \ldots, x_s$ are s concomitant observations besides the yield y then the composition of y may be written, assuming linear regression on the concomitant variates as

$$y=\beta+\tau+\eta_i+\rho_1x_1+\ldots+\rho_sx_s+\varepsilon' \qquad \ldots (7\cdot3)$$

The parameter τ in (7.3) may be called the effect of the variety corrected for concomitant variation. It is clear from these definitions of effects that these are deviations from a hypothetical standard so that the superiority of one treatment (or variety) over the other can be gauzed by a comparison of the effects alone. If, for any treatment of a given set of treatments, variations in the comcomitant variates alone produce significant variations in the yield then the comparison of the treatments must involve the elimination of variations of yield due to differences in the values of concomitant variates alone.

If the position of the plot for which the given variety is to be tried is determined at random from the plots of a block then the observed yield y becomes a stochastic variate in which case for the observed set of concomitant variates $x_1,x_2,\ldots,x_s$

$$\begin{aligned}\mathrm{E}(y)&=\beta+\tau+\mathrm{E}(\eta_i)+\rho_1x_1+\ldots+p_sx_s+\mathrm{E}(\varepsilon')\\ &\beta+\tau+\rho_1x_1+\ldots+\rho_sx_s\end{aligned} \qquad \ldots (7\cdot4)$$

for

$$\mathrm{E}(\eta_i)=(\Sigma\eta_i)/k=0 \text{ and } \mathrm{E}(\varepsilon')=0$$

If $y_1,y_2,\ldots,y_k$ are observations on k varieties tried on k plots with *random determination* of position for the varieties and $x_{1i},x_{2i},\ldots,x_{si}$ are the values of the s concomitant variates corresponding to the i-th variate then

$$\begin{aligned}\mathrm{E}(y_i)&=\beta+\tau_i+\rho_1x_{1i}\ldots+\rho_sx_{si}\\ i&=1, 2, \ldots, k\end{aligned} \qquad \ldots (7\cdot5)$$

constitute the observational equations corresponding to the k observed yields. The variables y_i are evidently correlated with equal correlation for any two and have equal variance. From the set of equations (7.5) we can construct k other equations.

$$\begin{aligned}\mathrm{B}&=(\Sigma y_i)/\sqrt{k},\ \mathrm{R}_i=\Sigma l_{ij}y_j\\ i&=1, 2, \ldots,(k-1)\end{aligned} \qquad \ldots (7\cdot6)$$

such that

$$\sum_j l_{ij}=0 \text{ and } \sum_j l_{ij}l_{rj}=0,\ \sum_j l^2_{ij}=1 \qquad \ldots (7\cdot7)$$

with the consequent relations

$$\mathrm{E(B)}=(\Sigma\tau)/\sqrt{k}+\beta\sqrt{k}+\rho_1(\Sigma x_{1i})/\sqrt{k}+\ldots \qquad \ldots (7\cdot8)$$

$$\begin{aligned}\mathrm{E}(\mathrm{R}_i)&=\Sigma\, l_{ij}\,\tau_j+\rho_1\Sigma l_{ij}x_{1j}+\ldots\\ i&=1, 2, \ldots,(k-1)\end{aligned} \qquad \ldots (7\cdot9)$$

In this scheme all the variables are uncorrelated but B and R_i are subject to different variances. The variance to which R_i is subject may be called the *intra block error* denoted by σ^2_{ia}. For all possible random arrangements the variation in B is caused by ε, the technical error, alone which may be represented by σ^2_ε.

A design may be defined as an arrangement of v kinds of elements, in b sets of $k_1, \ldots, k_b$ elements such that i-th kind of element is used r_i times and the i-th and j-th kind of elements occur together in λ_{ij} sets. Such arrangements when available can be used in actual experimentation which involves the testing of v varieties, or more generally treatments in the case of factorial experiments, by choosing the sets as blocks. If we assume that the intrablock error is of the same magnitude irrespective of the fertility of the block (in which case it is intrinsically connected with size and number of plots in a block) then it is theoretically advantageous to choose blocks of equal size so that all comparisons from the experiment may be of equal weight. The elements are identified with v treatments and the treatments (elements) belonging to a set are assigned at random with in a block. We discuss below only as an illustration the particular case when there are b blocks with k plots and each of the v treatments is used r times. We may set down the observational equations as.

$$E(B_i) = \beta_i/\sqrt{k} + (\Sigma\tau)/\sqrt{k} + \cdots \qquad \ldots \quad (7\cdot10)$$

$$i = 1, 2, \ldots, b$$

corresponding to the observed block totals, β_i being the value of β for the i-th block and $\Sigma\tau$ is the sum of treatment effects from the i-th block and b sets of $(k-1)$ equations of the type (7.9) from each block giving altogether bk equations. Estimation and testing of hypothesis is concerned only with parametric functions of τ's so that all linear combinations of observations involving the block totals are not permissable. Hence we may take the $b(k-1)$ equations of the type (7.9) alone as observational equations. This can also be observed after forming the normal equations. We thus arrive at the problem of estimation and testing of hypothesis involving $b(k-1)$ independent stochastic variates with equal variance and having their expectations as linear functions of parameters.

Denoting the effect of the i-th variety by τ_i and the regression parameters by $\rho_1, \rho_2, \ldots, \rho_s$ the normal equations leading to the best estimates are (observing that under conditions (7.7) $\sum_i l_{ij}^2 = (k-1)/k$ and $\sum_i l_{ij}\, l_{ir} = -1/k$)

$$Q_i(y) = \frac{r(k-1)}{k}\, t_i - \sum_j \frac{\lambda_{ij}}{k}\, t_j - \sum_p r_p Q_i(x_p) \qquad \ldots \quad (7\cdot11)$$

and

$$S(yx_i) - \frac{S[y(b_j)\, x_i(b_j)]}{k} \qquad \ldots \quad (7\cdot12)$$

$$= \Sigma t_p Q_p(x_i) + \Sigma\ r_p \left\{ \{ S(x_p x_i) - \frac{S[x_p(b_j) x_i(b_j)]}{k} \right\}$$

$$i = 1, 2, \ldots, s$$

where $Q_i(z)$ = sum of the observations on the z variate for the i-th treatment minus the means of the z variates for blocks in which the i-th treatment occurs, $S(u\, v)$=sum of products of the variables u and v, $w(b_s)$=total of the w variates from the s-th block. Observing that, as discussed earlier, the solutions t_i are given by

$$t_i = t_i^{(o)} - \Sigma r_p t_i^{(p)} \qquad \ldots \quad (7\cdot13)$$

where $t_i^{(o)}$ are the solutions of

$$Q_i(y) = r\frac{k-1}{k} t_i^{(o)} - \sum_j \frac{\lambda_{ij}}{k} t_j^{(o)} \qquad .. \quad (7{\cdot}14)$$

$$i=1, 2, \ldots, v$$

and $t_i^{(m)}$ are the solutions of

$$Q_i(x_m) = r\frac{k-1}{k} t_i^{(m)} - \Sigma \frac{\lambda_{ij}}{k} t_j^{(m)} \qquad .. \quad (7{\cdot}15)$$

and substituting for t_i its expression $t_i^{(o)} - \Sigma r_p t_i^{(p)}$ in (7 ·12) we get

$$T_{oi} - V_{oi} = \Sigma r_p (T_{pi} - V_{pi}) \qquad .. \quad (7{\cdot}16)$$

$$i=1, 2, \ldots, s$$

where

$$V_{ij} = \Sigma\, t_p^{(i)} Q_p(x_j) = \Sigma t_p^{(j)} Q_p(x_i) \qquad .. \quad (7{\cdot}17)$$

$$\left.\begin{aligned} T_{oi} &= S(yx_i) - \frac{S[y(b_p)x_i(b_p)]}{k} \\ T_{ij} &= S(x_i x_j) - \frac{S[x_i(b_p)x_j(b_p)]}{k} \end{aligned}\right\} \qquad .. \quad (7.18)$$

Defining $E_{oi} = T_{oi} - V_{oi}$, $E_{ij} = T_{ij} - V_{ij}$ which may be called the error sum of squares and products, we get the equations giving the regression parameters as

$$E_{oi} = \Sigma r_p E_{pi} \quad (i=1, 2, \ldots, s) \qquad .. \quad (7{\cdot}19)$$

To get the estimates of t_i we use the formula (7 ·13) with the values of regression parameters as obtained from (7 ·19). As shown earlier

$$V(t_i) = V(t_i^{(o)}) + V(\Sigma r_p t_i^{(p)}) \qquad (7{\cdot}20)$$

$$\text{Cov}\,(t_i t_j) = \text{Cov}\,(t_i^{(o)} t_j^{(o)}) + \text{Cov}\,(\Sigma r_p t_i^{(p)})\,(\Sigma r_p t_j^{(p)}) \qquad .. \quad (2{\cdot}21)$$

for each of which the expressions can be easily deduced following the discussion in the previous Sections.

To obtain the generalised variance due to error we take the minimised sum of squares

$$\Sigma[S(ly) - S(lt) - \rho_1 S(lx_1) - \ldots]^2 \qquad .. \quad (7{\cdot}22)$$

which on reduction comes out as

$$S(y^2) - \frac{S[y(b_p)^2]}{k} - \Sigma t_i^{(o)} Q_i(y)$$

$$-\Sigma r_p \left\{ S(yx_p) - \frac{S[y(b_i)x_p(b_j)]}{k} - \Sigma t^{(p)}{}_m Q_m(y) \right\}$$

$$= E_{oo} - \Sigma r_p E_{po} \qquad .. \quad (7{\cdot}23)$$

Let the degrees of freedom associated with this be d equal to b (k -1) minus the number of estimable independent parametric functions in τ's and ρ's. The generalised variance appropriate to test the hypothesis whether certain parametric functions have assigned values we minimise (7 ·22) subject to the restrictions imposed by the hypothesis and subtract the generalised variance due to error, the degrees of freedom associated with it being equal to the number of independent parametric functions whose values are assigned.

Thus, if the hypothesis to bs tested is H=TF′ where the rank of F is r we have to minimise (7 ·22) subject to H=TF′ and construct the generalised variance statistic by subtracting (7 ·23) from it. The formula (7 ·23) and the previous analysis suggest that this minimum can be calculated with the help of the residual vectors. In this case we find the least square solutions of

$$\Sigma[S(ly)-S(l\tau)]^2 \quad .. \quad (7 \cdot 24)$$

and

$$\Sigma[S(lx_i)-S(l\tau)]^2, \qquad i=1, 2, \ldots, s \quad .. \quad (7 \cdot 25)$$

subject to H=TF′ and calculate the residuals by substituting the solutions t's for τ's. Denoting the residuals by $[S(ly)-S(lt)]$ and $[S(lx_i)-S(lt)]$ where t stands for the corresponding solutions we can construct the following sum of squares and products.

$$E'_{oo}=\Sigma[S(ly)-S(lt)]^2=\Sigma S(ly)\,[S(ly)-S(lt)] \quad .. \quad (7 \cdot 26)$$

$$E'_{po}=\Sigma[S(ly)-S(lt)]\,[S(lx_p)-S(lt)] \quad .. \quad (7 \cdot 27)$$

$$=\Sigma S(ly)\,[S(lx_p)-S(lt)] \quad .. \quad (2 \cdot 28)$$

$$E'_{pi}=\Sigma S(lx_p)\,[S(lx_i)-S(lt)]$$

The minimum value is given by $E'_{oo}-\Sigma r'_p E'_{po}$ where r'_p are solutions of $E'_{oi}-\Sigma r'_p\, E'_{pi}=0$. In the case of general observational equations E(Y)=TA′+RX′, of which the above example is a complicated case, we may state the rule as follows.

Lemma 7a. *The minimum value of* $(Y-TA'-RX')(Y'-AT'-XR')$ *subject to* $H=TF'$ *is given by* $\eta\eta'-\Sigma r_i\xi_i\eta'$ *where r's are any solutions of* $\xi_i\bar{\eta}'=\Sigma r_p\xi_i\xi_p (i=1,2,\ldots,s)$ *and* η *and* ξ_i *are residual vectors obtained from the observational equations* $E(Y)=TA'$ *and* $E(X_i)=TA'$ *subject to the restristions* $H=TF'$.

As a particular case we may consider the testing of the hypothesis $\tau_1=\tau_2= \ldots \tau_v$. In this case the normal equations become independent of τ's. The minimised sum of squares is

$$\Sigma\,[S(ly)-\rho_1 S(lx_1)-\ldots]^2 \quad .. \quad (7 \cdot 29)$$

$$=T_{oo}-\Sigma r'_p T_{po}$$

where r_p' are solutions of

$$T_{oi}=\Sigma r'_p\, T_{pi} \quad .. \quad (7 \cdot 30)$$

where T_{ij}'s are as defined in (7 ·18). The generalised variance due to (v−1) independent parametric functions whose values are assigned is given by (7 ·29)−(7 ·23). The test is supplied by the statistic d[(7 ·29)−(7 ·23)]/(v−1)(7 ·23) which is distributed as the variance ratio with (v−1) and d degrees of freedom. The application to special cases is reserved for a subsequent communication.

We thus arrive at a scheme of computation for the covariance technique. The expressions that are to be evaluated require a logical subdivision of the arithmetical discussion into two sections, *the analysis of variance* which alone is sufficient for tests of significance of varietal effects when the regression parameters are not of importance and the *analysis of covariance* which supplies the necessary correction terms leading to precise tests of significance when the regression parameters are significant. An additional advantage to be gained in this is that they give us in a simple manner not only the systematic scheme of computation which can be carried out merely as a routine but also the structure of the experiment itself.

8. COMBINATION OF WEIGHTED OBSERVATIONS

The general theory of least squares as discussed in Section 6 can be extended to the general case of observations having unequal variances but known proportions. In this case the variances $\sigma^2_1, \ldots, \sigma_n^2$ of the observations $y_1, \ldots, y_n$ can be expressed as $\omega_1^2\sigma^2, \ldots, \omega_n^2\sigma^2$

where $\omega_1, \ldots, \omega_n$ are known quantities and σ is an unknown parameter. Let W denote a Kronecker matrix with $1/\omega_1, \ldots, 1/\omega_n$ as diagonal elements. If from Y we construct the vector Z by the transformation $Z = YW$ then the observational equations $E(Y) = TA'$ transform to $E(Z) = E(YW) = TA'W$ and the dispersion matrix of Z becomes $\sigma^2 I$ where σ is the unknown parameter defined above.

This problem can thus be reduced to the case already discussed. Instead of the variate y_i we consider the variate $z_i = y_i/\omega_i$ and in its expectation we divide the corresponding coefficients of the τ parameters by ω_i and proceed to the problem of estimation and testing of hypothesis with new variates and new matrix for the observational equations, the rank of which remains the same. The theory of weighted least squares can thus be reduced to the simple theory of least squares.

As an example, we may consider the following problem which occurs in the theory of regression. For a given value x_i of the concomitant variate x the stochastic variate y has the mean value βx_i and variance $\sigma^2 x_i$ so that the coefficient of variation of y is constant for all x arrays. For estimation and testing of hypothesis concerning the parameter β from pairs of observations $(x_1, y_1), (x_2\, y_2) \ldots (x_n y_n)$, we proceed as follows. If $E(y_i) = \beta x_i$ are observational equations then they transform to $E(z_i) = \beta \; (i = 1, 2, \ldots, n)$ where $z_i = y_i/x_i$. The normal equation leading to the best estimate of β is $\Sigma z_i = n\beta$, so that the estimate of β and its variance are $b = \Sigma z_i/n$ and σ^2/n respectively where σ^2 is estimated from the least squares $\Sigma(z_i - b)^2$ with $(n-1)$ degrees of freedom. To test the hypothesis that the value of $\beta = \beta_o$. we use the statistic

$$t = (b - \beta_o)\sqrt{n} \;/\; \sqrt{\Sigma(z_i - z)^2 / (n-1)} \qquad \ldots \quad (8 \cdot 1)$$

and refer it to the t distribution with $(n-1)$ degrees of freedom.

This treatment of linear regression may find an application in some biological and other problems where the mean value increases in a linear relation with the concomitant variate and the variance increases in such a manner that the coefficient of variation remains steady. As for instance the weight of dry paddy is linearly related to the weight of green paddy and the variance increases with increase in the green weight. The constancy of coefficient of variation is a plausible hypothesis in which case the above estimate of the regression coefficient is the most efficient and any hypothesis concerning the conversion factor may be tested as above.

9. THE THEORY OF STATISTICAL REGRESSION.

In this section some tests of significance connected with the regression coefficients are derived with the help of the results derived in Sections 4,5 and 6. The observations on the stochastic vector $Y = (y_1 \ldots y_n)$ are taken to be distributed normally and independently about their expectations with a common variance σ^2. The expectation vector is $E(Y) = TA' + RX'$ where A, R and X are as defined before, with some restrictions on the τ parameters. Following the method of analysis outlined in Section 5 we can write down the equations (5 ·5) giving the regression parameters as

$$\Sigma\, \eta_i \xi_{ji} = \rho_1 \Sigma \xi_{1i} \xi_{ji} + \ldots + \rho_k \,\Sigma\, \xi_{ki} \xi_{ji} \qquad \ldots \quad (9 \cdot 1)$$
$$j = 1, 2, \ldots, k$$

where the residual vectors ξ, η and the methods of calculation are given in (5 ·6) to (5 ·9). Due to lemma 7a we get the residual sum of squares when minimised with the whole set of parameters τ's and ρ's as given by $\eta\eta' - \Sigma r_i \xi_i \eta'$ where r_i's are the solutions of (9·1). The degrees of freedom associated with this can be obtained with the help of theorem 6a or theorem 6b. Let this be f.

The hypotheses to be tested in the theory of regression are certain linear hypotheses concerning the ρ parameters. If in particular the hypothesis is $\rho_i = 0$, then it will be testing the significance of the partial regression coefficient of the stochastic variate on the i-th concomitant variate. If the hypothesis to be testsd is $\rho_1 = \rho_2 = \ldots\ldots = \rho_k = 0$ (i.e. all ρ parameters are zeroes) it will be testing the significance of the multiple correlation coefficient.

Two methods are open to us. If the linear functions of the ρ parameters are estimable from the set of equations (9·1) then from their best estimates and variances and covariances, we can build up the appropriate generalised variance. If $P_1, P_2 \ldots . P_s$ ars the estimates of s independent linear functions with assigned values $p_1, p_2 \ldots p_s$ and $\sigma^2 d_{ij}$ is the covariance between P_i and P_j then the generalised variance with s degrees of freedom appropriate to P's is .. (9·2)

$$\sigma^2 V = \Sigma\Sigma\, d^{ij}\, (P_i - p_i)\, (P_j - p_j)$$

where d^{ij} are the elements of the matrix reciprocal to $((d_{ij}))$. If we denote the residual sum of squares by $\sigma^2 V_o$ then to test the present hypothesis we use the statistic $f\,V/sV_o$ as the F-statistic with s and f degress of freedom.

The alternative method is to find the residual sum of squares subject to the additional restrictions imposed by the hypothesis and get V by subtracting V_o from the above residual sum of squares. In the particular case when the hypothesis is $\rho_1 = \rho_2 = \ldots = \rho_k = 0$ we immediately observe that $\eta\eta'$ is the residual sum of squares under this hypothesis. Subtracting the error residual sum of squares from this we get $\Sigma r_i \xi_i \eta'$ as the appropriate generalised variance with s degrees of freedom. Hence to test the significance of the multiple correlation between the stochastic variate and the totality of the concomitant variates we use the statistic $f \Sigma r_i \xi_i \eta' / k[\eta\eta' - \Sigma r_i \xi_i \eta']$ as the F statistic with k and f degrees of freedom.

If s linear functions of the regression parameters have some assigned values, then to find the residual we have to minimise with these extra restrictions. If these restrictions are

$$\begin{array}{c} c_{11}\rho_1 + \ldots, + c_{1k}\rho_k = d_1 \\ \ldots . \qquad \ldots . \\ c_{s1}\rho_1 + \ldots, + c_{sk}\rho_k = d_k \end{array} \qquad .. \quad (9.3)$$

then introducing Lagrangian multiplies $\lambda_1, \ldots, \lambda_s$ we get the normal equations leading to the estimates of ρ's as .. (9,4)

$$\Sigma\, \eta_i\, \xi_{ji} = \rho_1\, \Sigma\, \xi_{1i}\xi_{ji} + \ldots + \rho_k \Sigma \xi_{ki}\xi_{ji} + \lambda_1 c_{11} + \ldots + \lambda_s c_{si}$$

$$j = 1,\, 2,\, \ldots k$$

and .. (9.5)

$$d_i = c_{i1}\rho_1 + \ldots + c_{ik}\rho_k$$

If $r_1', \ldots, r_k'$ is a solution to the above equations then the residual sum of squares is given by .. (9.6)

$$\Sigma\, (\eta_i - r'_1 \xi_{1i} - r'_2 \xi_{2i} - \ldots)^2$$
$$= \Sigma \eta_i^2 - \Sigma r'_i\, (2\Sigma \eta_j \xi_{ij} - \Sigma r'_s \Sigma \xi_{sk} \xi_{ik})$$

The generalised variance appropriate to these s functions is obtained by subtracting the error residual sum of squares from this. If this quantity be denoted by $\sigma^2 V$ then the statistic fV/sV_o which is distributed as the F statistic with s and f degrees of freedom can be used to test the above hypothesis. If the hypothesis is $\rho_i=0$ then in the equations (9 ·4) all λ's and ρ_i become zeroes. If $r_1',\ldots,r_k'$ are the solutions then the sum of squares appropriate to this is given by $\Sigma r_s'\xi_s\eta'-\Sigma r_i\xi_i\eta'$ with 1 degrees of freedom.

The distribution of the statistics used above on the null-hypothesis involve only the degrees of freedom and no other nuisance parameters so that tests of significance can be carried with exactitude. Because of this fact the distribution of these statistics is independent of the distribution of the concomitant variates so that the same tests of significance hold good independently of the distribution of the concomitant variates provided the relative distribution of the stochastic variates for given sets of concomitant variates are normal.

REFERENCES.

1. Bose, R.C. (1944). The fundamental theorem of linear estimation *Proc. Ind. Sc. Cong.* 4(3).
2. David, F. N. and Neyman, J. (1939). Extension of Markoff's theorem on least squares. *Biom. Res. Mem.* 2,238-249
3. Fisher, R. A. (1922). The goodness of fit of regression formula and the distribution of regression coefficients. *J. R. S. S.* 85,597 612
4. ————(1925), Application of Student's distribution. *Metron* 3, 30-104
5. Kolodziejczyk, St. (1935). On an important class of statistical hypotheses. *Biom.* 27, 161-190.
6. Markoff, A. A. (1904). Calculus of Probability. Russian Edition.
7. Neyman, J. and Pearson, E.S. (1930). On the problem of two samples. *Bull de Ac. pol. des Sc et des let.* 73-96.
8. ————(1931). On the problem of k samples. *ibid* 460-482.
9. ————(1933). On the problem of most efficient tests of statistical hypotheses. *Phil. Trans. of Roy. Sec.* 231A, 289-337.
10. Rao, C.R. (1943). Thesis submitted to the Calcutta University for the M.A. degree.
11. ————(1944). On the linear set up leading to intra and interblock informations. *Science and Culture* 10,259-260.
12. ————(1945a). Generalisation of Markoffs' theorem and tets of linear hypotheses. *Sankhya* 7(1), 9-16.
13. ————(1945b). Markoff's theorem with linear restrictions on parameters. *Sankhya* 7(1), 16-19.
14. ————(1945c). Studentised tests of linear hypotheses. *Science and culture* 11.
15. "Student" (1908). The probable error of the mean. *Biom.* 6, 1-25.

Paper received 7 December, 1944.

THEORIES OF PROBABILITY

By JAGJIT SINGH

Traffic Officer, E.I.Railway

THE CLASSICAL AND VON MISES' THEORIES OF PROBABILITY

As is well known there is no universally accepted definition of Probability. Four different definitions have been attempted.

(i) *Bernoulli-Laplace Definition.* If under certain conditions there are n exhaustive, mutually exclusive and equally likely cases and m of them are favourable to an event A, then the probability of the event A is defined as the ratio m/n.

The main objection to this definition is that it can be applied only in very simple and comparatively unimportant cases like games of chance. It is of no use in dealing with problems of insurance, statistics and theory of errors etc. Besides the definition involving as it does the concept of 'equally likely cases,' is circular. Certain writers try to avoid appealing to this concept by defining probability as follows :

If under certain conditions there are n possible cases and m of them are favourable to an event A, then the probability of A is the ratio m/n.

This avoids reference to "equally likely cases" but lands us in fresh difficulties. Consider, for instance, the following example quoted by Jeffreys. There are two boxes, one containing one black and one white ball, and the other one white and two black balls. A box is selected at random and then a ball is drawn at random from that box. What is the probability that a white ball will be drawn ? There are 5 balls and any one of them may be selected. There are thus 5 possible cases of which two will lead to the selection of a white ball. The probability of selecting a white ball according to the above definition should therefore be $(\frac{2}{5})$but all authorities including those who uphold this definition would give $(\frac{1}{2})(\frac{1}{2})+(\frac{1}{2})(\frac{1}{3})=(\frac{5}{12})$ as the probability. The fallacy, of course, lies in the fact that 5 mutually exclusive cases that are possible are not "equally likely" and therefore it is not permissible to take the ratio $(\frac{2}{5})$ as the probability of picking up a white ball. It is easy to give other examples in which the definition would lead to values of no use whatsoever e.g. according to this definition the probability of throwing 2 in the case of a dice is $(\frac{1}{6})$ irrespective of the fact whether the dice is biased or not.

(ii) *Von Mises' Definition.* The logical foundations of Bernoulli-Laplace theory were first rigorously examined by Venn who defined probability as lim (s/n)as $n \rightarrow \infty$ where s is the number of times the event A occurs in a series of n trials. Von Mises has taken up this idea and maintains that the notion of probability is applicable only to events capable of infinite repetition. Von Mises thus considers a collective, an unlimited sequence K of similar events. His first postulate is that in such a collective, (s/n) tends to a definite limit p, when n tends to ∞, where s is the number of times event A occurs in n trials and p is then defined to be the probability of the event A. Von Mises' second postulate—and here he goes a step beyond Venn,the originator of the frequency notion of probability—is that the limit p is unaffected by what he calls "a place selection". There are two objections to Von Mises' definition, (a) The existence of limit in the collective denies the possibility of complete randomness. In fact some authorities consider the concept of a collective self-contradictory. (*b*) The second postulate of Von Mises regarding the limit being unaffected by a place selection has not been given a precise mathematical meaning. We shall discuss these objections in greater detail in the sequel. Meanwhile we call attention to the fact that difficulties regarding expressing the second postulate of Von Mises in precise mathematical language has prompted an axiomatic approach.

(iii) Kolmogoroff, Cramer and others define probability of an event as a definite number associated with that event and then lay down certain axioms expressing the fundamental rules for operations with such numbers, thus making the theory of probability "a part of the theory of additive set functions". Objections have been raised against this theory too. In the first place it is urged that an axiomatic theory of Probability is of no more utility than a similar theory of Electromagnetism or Gravitation. A theory of probability must in all circumstances remain a theory of certain observable phenomena. Secondly probability is defined as the ratio of the measures of sets of points and the measure of a continuous set is ambiguous unless it is separately defined.

(iv) An entirely different approach is that of Jeffreys' who has given an elaborate mathematical development of an idea of Keynes viz. probability is the degree of reasonable belief which a proposition merits on given evidence. Jeffreys' fundamental hypothesis is that every proposition p on data q has a unique numerical probability and that probabilities of different propositions on different data are comparable. But he gives no indication how this number is to be assessed in any given case. Majority of the statisticians and mathematicians have therefore rejected Jeffreys' approach. There are a number of other objections though they are less serious.

Jeffreys' proof of the product rule holds good only in the particular case when Q the disjunction, is resolvable into a set of *equally* probable and exclusive alternatives on data p. (See Theorem 9 page 24 of his Theory of Probability). In cases where Q cannot be analysed into a set of equally probable and exclusive alternatives, the validity of the product is simply assumed. On the other hand it is possible to prove rigorously the product rule on the basis of any of the definitions quoted above.

Secondly Jeffreys' proof of the theorem, very vital for his theory, that probability of a proposition q, on previous information H and two batches of data p_1 and p_2 is the same in whatever order the data are taken is faulty. (See his Theory of Probability page 35). What is required to be proved is that $P(q|p_1\,p_2H)=P(q|p_2p_1H)$ which is equal to the probability of q, when p_1, p_2 and H are taken into account simultaneously. In his proof, Jeffreys freely replaces terms like $P(q|p_1p_2H)$ by $P(q|p_2p_1H)$ and terms like $P(p_2|p_1qH)$ by $P(p_2|qp_1H)$ in the expressions occuring in the denominators of the terms of the third equation (article 1·5). This interchange entirely begs the question at issue.

2. A NEW THEORY OF PROBABILITY

Now what is Probability ? The question is still as difficult as when Poincare wrote his famous treatise over 50 years ago. In the first place it must be admitted with Venn and Von Mises that we can speak of probability of only such events as are capable of infinite repetition. Now if in a sequence of n trials an event A occurs s times, it is observed that the frequency ratio s/n exhibits a peculiar regularity. This regularity may be roughly described by saying that s/n tends to become more or less constant as n increases. Now Von Mises describes this regularity or stability of the frequency ratio by postulating that s/n

tends to a limit p as n tends to infinity. This means that given $\varepsilon>0$, however small, a number $n_0(\varepsilon)$ can be found such that

$$\left|\frac{s}{n}-p\right|<\varepsilon \qquad .. \quad (2{\cdot}1)$$

for all $n\geqslant n_0$

Now suppose we have found a number n_0 such that (2·1) holds for all $n\geqslant n_0$. Consider now m further trials of which say μ are "successes". Then

$$\left|\frac{s+\mu}{n_0+m}-p\right|<\varepsilon \qquad .. \quad (2{\cdot}2)$$

for all m, whatsoever. Now if (2·2) holds

$$p-\varepsilon<\frac{s+\mu}{n_0+m}<p+\varepsilon \qquad .. \quad (2{\cdot}3)$$

$$\mu<(p+\varepsilon)\ (m+n_0)-s \qquad .. \quad (2{\cdot}4)$$

But from (2·1) $\qquad (p-\varepsilon)n_0<s<(p+\varepsilon)n_0$

Therefore, from (2·4) $\qquad \mu<(p+\varepsilon)(m+n_0)-(p-\varepsilon)n_0$

Or, $\qquad \mu<(p+\varepsilon)m+2\varepsilon n_0$

Similarly, $\qquad \mu>(p-\varepsilon)m-2\varepsilon n_0$

Hence, the number of "successes" μ in m trials subsequent to n trials must lie between

$$(p-\varepsilon)\ m-2\varepsilon n_0 \quad \text{and} \quad (p+\varepsilon)\ m+2\varepsilon n_0 \qquad .. \quad (2{\cdot}5)$$

This shows that the existence of n_0 for a given ε imposes some restrictions on the sequence of Von Mises, viz., that after n_0 the number of successes occuring in the sequence are restricted to lie between two predetermined values. So long as m is small, the limits are wider than the a *priori* limits between which μ could lie viz. o and m, but when m increases the limits, (2·5) become much narrower than the a priori limits of μ viz. o to m. This may be illustrated by a numerical example. Suppose $p=\frac{1}{2}$, $n_0=10{,}000$, and $\varepsilon=\cdot001$, then the existence of a limit implies that

$$\left|\frac{s}{n}-\tfrac{1}{2}\right|<\cdot001, \text{ for all } n\geqslant 10{,}000$$

Now take $m(=10)$ further trials. Then a priori limits for μ are 0 to 10 but the existence of limit gives rise to limits (2·5) which in this instance are—15·01 and 25·01. These limits cover the *a priori* limits entirely and the inquality (2·2) will hold whatever the result of $(m=10)$ trials. But if $m=100$, the *a priori* limits for μ are 0 and 100 but limits (2·5) are 29·9 and 70·1, which do not cover the *a priori* limits. As m increases the limits (2·5) become smaller in comparison with *a priori* limits for μ

The existence of limit of the frequency ratio therefore involves a restriction on the randomness of the collective. Complete randomness and the existence of limit for the frequency ratio do not therefore go together and there is no way out.

In his book Probability, Statistics and Truth, Von Mises has tried to answer objections raised against the first condition imposed on a collective (page 126). Therein he explains that there is no inconsistency between his first condition and Bernoulli's theorem that the probability with which relative frequency approaches p tends to 1 as n tends to infinity. But the real objection to the postulate is as stated above viz. that if the limit exists, $|s/n-p|<\varepsilon$ should hold for *all* $n\geqslant n_0$

Definition of Quasi-limit

As an alternative to and an improvement over Von Mises' definition of probability, the following definition based upon the existence of, what I call, a quasi-limit is being introduced. Consider the infinite collective and the sequence (S_n) of its frequency ratios, S_n being equal to s/n, where s is the number of 'successes' in n trials. We shall say that S_n tends to a "quasi-limit" p as n tends to ∞ whenever the following conditions hold. We shall denote "quasi-limit" by q-limit as well.

Given two positive numbers ε, η as small as we please, then two numbers n_o and m_o can be found such that of the m terms $S_{n+1}, S_{n+2}, S_{n+3} \ldots S_{n+m}$, π terms satisfy the inequality

$$| S_{n+i} - p | < \varepsilon \qquad \ldots \quad (2\cdot6)$$

where $\pi/m > 1-\eta$ whenever $n \geqslant n_o$ and $m \geqslant m_o$, the value of m_o depending upon ε, η and n.

In this manner provision is made for a progressively diminishing fraction of the terms S_{n+i}, for $i=0, 1, \ldots\ldots m$, not satisfying the above inequality. The empirical law of the regularity of frequency ratios referred to above is thus mathematically expressed by the postulate that sequence (S_n) tends to "quasi-limit" p as n tends to ∞.

We shall now prove that if a sequence tends to a "quasi-limit", the "quasi-limit" must be unique.

For if not, let there be two quasi-limits p and p' of a sequence (S_n). Let $p' > p$. Then for any given $\varepsilon < (p'-p)/2$ and any η as small as we like, we can find numbers n_o, m_o, and n'_o, and m'_o, such that. $| S_{n+i} - p | < \varepsilon$ for π values of $i = 1, 2, \ldots\ m$ where $\pi/m > 1-\eta$ for all $n \geqslant n_o$ and $m \geqslant m_o$ and also $|S_{n+i} - p'| < \varepsilon$ for π' values of $i = 1, 2, \ldots m'$ where $\pi'/m_1 > 1-\eta$ for all $n \geqslant n'_o$ and $m' \geqslant m'_o$.

Let us take any n greater than greater of the two numbers n_o and n'_o, and m greater than the greater of the two numbers m_o and m'_o.

Then consider the inequalities

$$|S_{n+i} - p | < \varepsilon \text{ and } |S_{n+i} - p'| < \varepsilon \qquad \ldots \quad (2\cdot7)$$

Inequality (2 ·6) holds for π terms out of the m terms $S_{n+1}, S_{n+2} \ldots S_{n+i}, \ldots, S_{n+m}$ where $\pi > (1-\eta)m$. Thus the number of the terms of the sequence $S_{n+1}, S_{n+2}, \ldots S_{n+m}$ which lie within the range $(p-\varepsilon, p+\varepsilon)$ exceeds $(1-\eta)\, m$. Hence the number of terms of this sequence (of m terms) which lie outside $p \pm \varepsilon$ cannot exceed $m\eta$. Now only some of the terms of this sequence which lie outside $p \pm \varepsilon$ can possibly lie within the range $p' \pm \varepsilon$ for any term lying within $p \pm \varepsilon$ clearly lies outside the range $p' \pm \varepsilon$. Hence, $m\eta$ is greater than the number of terms of the sequence $S_{n+1}, \ldots S_{n+i}, \ldots S_{n+m}$ lying outside $p \pm \varepsilon$ and therefore also within $p' \pm \varepsilon$

But the number of terms of the sequence $S_{n+1}, \ldots S_{n+i}, \ldots S_{n+m}$ lying within $p' \pm \varepsilon$ exceeds $m\ (1-\eta)$ by hypothesis.

Therefore, $m\eta > m(1-\eta)$ or $\eta > 1/2$ which contradicts the hypothesis that η is an arbitrarily small number. Hence $p = p'$.

The addition and product rule of probabilities can also be easily proved on the basis of the quasi-limit definition of probability given above.

Addition Rule. Consider a collective of events with several mutually exclusive outcomes A, B, ... Let s/n be the frequency ratio of A, and s'/n that of B in a series of n trials. Then the frequency ratio for the event A or B is clearly $(s+s')/n$.

We have to prove that if s/n tends to quasi-limit p and s'/n to quasi-limit p', then $(s+s')/n$ will tend to quasi-limit $p+p'$ when n tends to infinity. Let $S_n = s/n$, $S'_n = s'/n$ and $\Sigma_n = (s+s')/n$. By hypothesis given ε, η any two arbitrarily small positive numbers, n_0 and m_0 can be found such that $|S_{n+i} - p| < \varepsilon$ holds for π terms of $S_{n+1}, S_{n+2}, \ldots S_{n+m}$ where $\pi > (1-\eta)m$, where $n \geqslant n_0$ and $m \geqslant m_0$ And also $|S'_{n+i} - p'| < \varepsilon$ holds for π terms of $S'_{n+1}, \ldots$ $\ldots S'_{n+m}$, where $\pi' > (1-\eta)m$ for all $n \geqslant n'_0$ and $m \geqslant m'_0$ Let N_0 be greater of the two numbers d_0 and n'_0, and M_0 be greater of the numbers m_0, m'_0. Then for $n > N_0$ and $m > M_0$

$$\begin{aligned} |\Sigma_{n+i} - (p+p')| &= |S_{n+i} - p + S'_{n+i} - p'| \\ &< |S_{n+i} - p| + |S'_{n+i} - p'| < 2\varepsilon \end{aligned} \quad .. \quad (2{\cdot}8)$$

holds for θ terms of the sequence $\Sigma_{n+1}, \Sigma_{n+2}, \ldots \Sigma_{n+m}$ where $\theta > (1-2\eta)m$

For the number of terms of $S_{n+1}, \ldots S_{n+i}, \ldots S_{n+m}$ for which $|S_{n+i} - p| < \varepsilon$ does not hold is at most $m\eta$ and of those for which $|S'_{n+i} - p'| < \varepsilon$ does not hold is at most $m\eta$. Hence the number of terms for which one or the other or both the inequalities do not hold is at most $2m\eta$.

Product Rule. Consider a collective of events with several mutually exclusive outcomes A, B, ...and a second collective of events with mutually exclusive outcomes A′, B′, ... Now by associating the result of ith trial in the 1st collective with that of ith trial in the second we can construct a new collective of compound events with several mutually exclusive outcomes (A, A″), (A,B′) etc. etc. Let s/n be the frequency ratio of A in a series of n trials of the first collective ; then quasi-limit of s/n exists and is the probability p of the event A. Let μ/n be the frequency ratio of the compound event (A,A′) *in n trials of the compound collective* formed by associating the two collectives. Let us now pick up those items of the compound collective in which the first event is A. There can be only s such items. Of these s items there will obviously be only μ in which the second event is A′. We assume that quasi-limit of μ/s exists and is q. We will now prove that quasi-limit of μ/s exists and is pq. In other words the probability of the compound event (A, A′) = probability of A × probability of A′ given that A has already occurred. Or, $P(A,A') = P(A) \times P(A'|A)$. Since quasi-limit of $s/n (= S_n)$ exists we can find n_0 and m_0 for any $\varepsilon, \eta > 0$ such that $|S_{n+i} - p| < \varepsilon$ for π values of $i = 1, 2, \ldots m$ where $\pi > m(1-\eta)$ whenever $n \geqslant n_0$ and $m \geqslant m_0$.

Similarly since quasi-limit of $\mu/s (= S'_s)$ exists and is q we can find s_0 and M_0 such that $|S'_{s+t} - q| < \varepsilon$ for ρ values of $t = 1, 2, \ldots$ M where $\rho > M(1-\eta)$ whenever $s \geqslant s_0$ and $M \geqslant M_0$. Let $\Sigma_n = \frac{\mu}{n} = \frac{\mu}{s} \cdot \frac{s}{n} = S'_s \cdot S_n$ Now take $n > n_0$ and *at the same time* so big that s the frequency of A in n trials $\geqslant s_0$. Also take m greater than m_0. Then, $|S_{n+i} - p| < \varepsilon$ holds for π terms of the sequence $S_{n+1}, \ldots S_{n+i}, \ldots S_{n+m}$ such that $\pi > (1-\eta)m$. Now suppose in $(n+m)$ trials first event, A occurs $s+\theta$ times. By taking m sufficiently large, θ can be made as large as we please. Let us take m so large that $\theta \geqslant M_0$. Hence for such m and n chosen as above $|S'_{s+t} - q| < \varepsilon$ holds for ρ terms of the sequence $S'_{s+1} \ldots S'_{s+t} \ldots S'_{s+\theta}$ where $\rho > (1-\eta)\theta$. Now $\Sigma_{n+i} = \frac{\xi}{n+i}$, ξ being the frequency of the compound event (A, A′) in $(n+i)$ trials.

Definition of Quasi-limit

As an alternative to and an improvement over Von Mises' definition of probability, the following definition based upon the existence of, what I call, a quasi-limit is being introduced. Consider the infinite collective and the sequence (S_n) of its frequency ratios, S_n being equal to s/n, where s is the number of 'successes' in n trials. We shall say that S_n tends to a "quasi-limit" p as n tends to ∞ whenever the following conditions hold. We shall denote "quasi-limit" by q-limit as well.

Given two positive numbers ε, η as small as we please, then two numbers n_0 and m_0 can be found such that of the m terms $S_{n+1}, S_{n+2}, S_{n+3} \ldots S_{n+m}$, π terms satisfy the inequality

$$| S_{n+i} - p | < \varepsilon \qquad .. \quad (2\cdot6)$$

where $\pi/m > 1-\eta$ whenever $n \geqslant n_0$ and $m \geqslant m_0$, the value of m_0 depending upon ε, η and n.

In this manner provision is made for a progressively diminishing fraction of the terms S_{n+i}, for $i=0, 1, \ldots\ldots m$, not satisfying the above inequality. The empirical law of the regularity of frequency ratios referred to above is thus mathematically expressed by the postulate that sequence (S_n) tends to "quasi-limit" p as n tends to ∞.

We shall now prove that if a sequence tends to a "quasi-limit", the "quasi-limit" must be unique.

For if not, let there be two quasi-limits p and p' of a sequence (S_n). Let $p' > p$. Then for any given $\varepsilon < (p'-p)/2$ and any η as small as we like, we can find numbers n_0, m_0, and n'_0, and m'_0, such that. $| S_{n+i} - p | < \varepsilon$ for π values of $i=1,2,\ldots\ m$ where $\pi/m > 1-\eta$ for all $n \geqslant n_0$ and $m \geqslant m_0$ and also $|S_{n+i} - p'| < \varepsilon$ for π' values of $i=1,2,\ldots m'$ where $\pi'/m_1 > 1-\eta$ for all $n \geqslant n'_0$ and $m' \geqslant m'_0$.

Let us take any n greater than greater of the two numbers n_0 and n'_0, and m greater than the greater of the two numbers m_0 and m'_0.

Then consider the inequalities

$$|S_{n+i} - p | < \varepsilon \text{ and } |S_{n+i} - p'| < \varepsilon \qquad .. \quad (2\cdot7)$$

Inequality (2 ·6) holds for π terms out of the m terms $S_{n+1}, S_{n+2} \ldots S_{n+i}, \ldots, S_{n+m}$ where $\pi > (1-\eta)m$. Thus the number of the terms of the sequence $S_{n+1}, S_{n+2}, \ldots S_{n+m}$ which lie within the range $(p-\varepsilon, p+\varepsilon)$ exceeds $(1-\eta)\, m$. Hence the number of terms of this sequence (of m terms) which lie outside $p \pm \varepsilon$ cannot exceed $m\eta$. Now only some of the terms of this sequence which lie outside $p \pm \varepsilon$ can possibly lie within the range $p' \pm \varepsilon$ for any term lying within $p \pm \varepsilon$ clearly lies outside the range $p' \pm \varepsilon$. Hence, $m\eta$ is greater than the number of terms of the sequence $S_{n+1}, \ldots S_{n+i}, \ldots S_{n+m}$ lying outside $p \pm \varepsilon$ and therefore also within $p' \pm \varepsilon$

But the number of terms of the sequence $S_{n+1}, \ldots S_{n+i}, \ldots S_{n+m}$ lying within $p' \pm \varepsilon$ exceeds $m\ (1-\eta)$ by hypothesis.

Therefore, $m\eta > m(1-\eta)$ or $\eta > 1/2$ which contradicts the hypothesis that η is an arbitrarily small number. Hence $p = p'$.

The addition and product rule of probabilities can also be easily proved on the basis of the quasi-limit definition of probability given above.

Addition Rule. Consider a collective of events with several mutually exclusive outcomes A, B, ... Let s/n be the frequency ratio of A, and s'/n that of B in a series of n trials. Then the frequency ratio for the event A or B is clearly $(s+s')/n$.

We have to prove that if s/n tends to quasi-limit p and s'/n to quasi-limit p', then $(s+s')/n$ will tend to quasi-limit $p+p'$ when n tends to infinity. Let $S_n=s/n$, $S'_n=s'/n$ and $\Sigma_n=(s+s')/n$. By hypothesis given ε, η any two arbitrarily small positive numbers, n_0 and m_0 can be found such that $|S_{n+i}-p|<\varepsilon$ holds for π terms of $S_{n+1}, S_{n+2}, \ldots S_{n+m}$ where $\pi>(1-\eta)m$, where $n \geqslant n_0$ and $m \geqslant m_0$ And also $|S'_{n+i}-p'|<\varepsilon$ holds for π terms of $S'_{n+1}, \ldots \ldots S'_{n+m}$, where $\pi'>(1-\eta)m$ for all $n \geqslant n'_0$ and $m \geqslant m'_0$ Let N_0 be greater of the two numbers d_0 and n'_0, and M_0 be greater of the numbers m_0, m'_0. Then for $n>N_0$ and $m>M_0$

$$\begin{aligned} |\Sigma_{n+i}-(p+p')| &= |S_{n+i}-p+S'_{n+i}-p'| \\ &< |S_{n+i}-p| + |S'_{n+i}-p'| < 2\varepsilon \end{aligned} \qquad .. \quad (2{\cdot}8)$$

holds for θ terms of the sequence $\Sigma_{n+1}\cdot\Sigma_{n+2}, \ldots \Sigma_{n+m}$ where $\theta>(1-2\eta)m$

For the number of terms of $S_{n+1}, \ldots S_{n+i}, \ldots S_{n+m}$ for which $|S_{n+i}-p|<\varepsilon$ does not hold is at most $m\eta$ and of those for which $|S'_{n+i}-p'|<\varepsilon$ does not hold is at most $m\eta$. Hence the number of terms for which one or the other or both the inequalities do not hold is at most $2m\eta$.

Product Rule. Consider a collective of events with several mutually exclusive outcomes A, B, ...and a second collective of events with mutually exclusive outcomes A′, B′, ... Now by associating the result of ith trial in the 1st collective with that of ith trial in the second we can construct a new collective of compound events with several mutually exclusive outcomes (A, A″), (A,B′) etc. etc. Let s/n be the frequency ratio of A in a series of n trials of the first collective; then quasi-limit of s/n exists and is the probability p of the event A. Let μ/n be the frequency ratio of the compound event (A,A′) in n trials of the compound collective formed by associating the two collectives. Let us now pick up those items of the compound collective in which the first event is A. There can be only s such items. Of these s items there will obviously be only μ in which the second event is A′. We assume that quasi-limit of μ/s exists and is q. We will now prove that quasi-limit of μ/s exists and is pq. In other words the probability of the compound event (A, A′)=probability of A×probability of A′ given that A has already occurred. Or, $P(A,A')=P(A)\times P(A'|A)$. Since quasi-limit of $s/n(=S_n)$ exists we can find n_0 and m_0 for any $\varepsilon, \eta>0$ such that $|S_{n+i}-p|<\varepsilon$ for π values of $i=1,2,\ldots m$ where $\pi>m(1-\eta)$ whenever $n \geqslant n_0$ and $m \geqslant m_0$.

Similarly since quasi-limit of $\mu/s(=S'_s)$ exists and is q we can find s_0 and M_0 such that $|S'_{s+t}-q|<\varepsilon$ for ρ values of $t=1,2,\ldots$ M where $\rho>M(1-\eta)$ whenever $s \geqslant s_0$ and $M \geqslant M_0$. Let $\Sigma_n=\frac{\mu}{n}=\frac{\mu}{s}\cdot\frac{s}{n}=S'_s \,.\, S_n$ Now take $n>n_0$ and *at the same time* so big that s the frequency of A in n trials $\geqslant s_0$. Also take m greater than m_0. Then, $|S_{n+i}-p|<\varepsilon$ holds for π terms of the sequence $S_{n+1}, \ldots S_{n+i}, \ldots S_{n+m}$ such that $\pi>(1-\eta)m$. Now suppose in $(n+m)$ trials first event, A occurs $s+\theta$ times. By taking m sufficiently large, θ can be made as large as we please. Let us take m so large that $\theta \geqslant M_0$. Hence for such m and n chosen as above $|S'_{s+i}-q|<\varepsilon$ holds for ρ terms of the sequence $S'_{s+1} \ldots S'_{s+i} \ldots S'_{s+\theta}$ where $\rho>(1-\eta)\theta$. Now $\Sigma_{n+i}=\dfrac{\xi}{n+i}$, ξ being the frequency of the compound event (A, A′) in $(n+i)$ trials.

Then if in $(n+i)$ trials the first event A occurs $s+t$ times,

$$\Sigma_{n+i}=\frac{\xi}{n+i}=\frac{\dot{\xi}}{s+t}\cdot\frac{s+t}{n+i}=S'_{s+t}\cdot S_{n+1} \qquad \text{.. (2·9)}$$

But $S_{n+i}=p\pm\delta_i$ (δ_i being a positive number<1) for π values of $i=1,2,\ldots m$. and $S'_{s+t}=q\pm\delta_t'\varepsilon$ (δ_t' being a positive number <1) for ρ values of $t=1,2,\ldots\ldots\theta$.

Now from the relation (2·9)

$$\begin{aligned} \mid \Sigma_{n+i}-pq \mid &= \mid S'_{s+t}\, S_{n+i}-pq \mid \\ &= \mid (q\pm\delta'_t\varepsilon)\,(p\pm\delta_i\varepsilon)-pq. \mid \qquad \text{.. (2·10)} \\ &< 2\varepsilon+\varepsilon^2=\omega \text{ say.} \end{aligned}$$

for all $\Sigma_{n+1}, \Sigma_{n+2} \ldots \Sigma_{n+i} \ldots \Sigma_{n+m}$ except at the most for $m\eta+\theta\eta=(m+\theta)\eta$ terms. But $\theta\leqslant m$, hence the above inequality holds for all $\Sigma_{n+1}, \Sigma_{n+2} \ldots \Sigma_{n+m}$ except at the most for $2m\eta$ terms, when $m\geqslant m_0$.

Therefore, given two arbitrarily small positive numbers ω and 2η we can find n_0 and m_0 such that $|\Sigma_{n+i}-pq|<\omega$ holds for π terms of the sequence $\Sigma_{n+1}\ldots\Sigma_{n+i}\ldots\Sigma_{n+m}$ where $\pi>(1-2\eta)m$ when $n\geqslant n_0$ and $m\geqslant m_0$

REFERENCES

1. JEFFREYS : Theory of Probability.
2. Von. MISES : Probability, Statistics and Truth
3. ———— Währscheinlichkeitsrechnung.
4. CRAMER. H. : Random Variables and Probabilty Distributions.
5. Kolmogoroff. A : Gründbegriffe der Wahrscheinlichkeitsrechnung

Paper received : 20 July 1945.

A STATISTICAL APPROACH TO THE PROBLEM OF CHRONOLOGY OF SHAKESPEARE'S PLAYS.

By M. R. YARDI, I.C.S.
Sind

In determining the time order of Shakespeare's plays, the main effort of recent scholarship has been to supplement external evidence by a close analysis of style and metre. A great deal of work has already been done on the application of metrical tests to this problem, but most of it is unsatisfactory. Professor König, for instance, has worked out the percentages of the metrical characteristics to the total number of blank-verse lines for each play, and then, by direct comparison, established several tests in his *Der Vers in Shakespeares Dramen*, 1888. Similar methods have been followed by Professor Conard and his disciples in their attempts to determine the intervals between the composition of different acts and passages in each play. Their methods, to say the least, will be hardly acceptable to anyone with the slightest knowledge of statistics, and even Professor Chambers, who describes them as statistical, is sceptical about their validity. So far as my knowledge goes, no one has yet attempted a statistical examination of the available metrical data. It is, therefore, proposed to show how satisfactory statistical methods can be applied to a problem of this kind and to indicate their limitations, if any.

I may explain at the outset that I have not bothered myself about the various theories of collaboration, interpolation, abridgement and revision, which have been advanced by some Shakespearean scholars on the slightest metrical grounds. I have taken the plays as a whole, except *The Taming of the Shrew* and *Henry VIII* for which I have used only those parts which are commonly ascribed to Shakespeare.

Again it has not been found possible to utilize the whole body of collected data. Most of it had to be rejected either as having no chronological value like the distribution of prose or as being too sparse to be of any statistical use like the light and weak endings, studied by Professor J. K. Ingram. Of the available metrical data, only the variations of stress and pause appeared to me to have some chronlogical significance. Here again we have no reliable figures for the distribution of "inverted stresses" and so I had to omit them from consideration altogether. The most easily distinguishable type of stress variation are the redundant final syllables pointed out by Charles Bathurst. The figures (col. 1, Table 1) for these are borrowed from Fleay, though his accuracy is not above doubt. As regards pausation, it is almost certain that Shakespeare's language became less and less formal with the growing complexity of his thoughts, and so in its last phase we discover not only a high proportion of overflows but also a marked tendency to end a speech in mid-line. Unfortunately enumeration of these is difficult, because (1) it is largely subjective and may differ with different persons, thus making verification impossible, and (2) it depends on the punctuation of modern editors, which may not faithfully represent Shakespeare's usage. But this will not matter so long as the enumerator is consistent and confines himself to strong pauses. Here too I have to content myself with the rough counts given by Professor Chambers (col. 2, 3, Table 1). Since these counts will obviously vary with the total number of speech lines in a play, this is also taken as a fourth variable (Fleay's figures, col. 4, Table 1). It is a pity that the statistical analysis has to be based on rough and inaccurate counts, but it cannot be helped.

TABLE 1. CHARACTERISTICS AND DISCRIMINANT SCORES

Name of the Play	x (1)	y (2)	z (3)	w (4)	discriminant scores (5)	(6)	(7)
1 Henry VI, 2	332	13	270	3162	·66	1·18	1·74
2 Henry VI, 3	366	8	340	2904	·49	·88	1·55
3 Henry VI, 1	191	17	233	2677	·81	1·18	2·00
4 Richard III	638	66	423	3619	1·38	1·99	2·16
5 Comedy of Errors	198	8	171	1777	·65	·92	1·27
6 Titus Andronicus	200	17	303	2523	·83	1·06	2·11
7 Taming of the Shrew	208	21	194	1443	1·12	1·30	1·64
8 Two Gentelmen of Verona	269	43	174	2292	1·32	1·75	1·96
9 Love's Labour's Lost	26	11	79	2785	·57	1·06	1·87
10 Romeo and Juliet	168	71	496	3050	1·31	1·35	3·43
11 Richard II	258	34	293	2757	1·11	1·46	2·30
12 Midsummer-Night's Dream	59	28	131	2174	1·08	1·39	2·17
13 King John	151	64	357	2570	1·37	1·50	2·96
14 Merchant of Venice	325	79	369	2658	1·51	1·77	2·66
15 Henry IV, 1	92	43	235	3176	1·09	1·49	2·78
16 Henry IV, 2	221	43	208	3446	1·10	1·69	2·46
17 Much Ado About Nothing	145	35	126	2825	1·10	1·62	2·20
18 Henry V	3363	31	219	3381	1·02	1·65	2·03
19 Julius Caesar	410	129	437	2477	1·77	1·96	2·78
20 As you like it	23	34	177	2856	1·11	1·62	2·10
21 Twelfth Night	167	44	161	2690	1·23	1·68	2·30
22 Hamlet	528	194	552	3929	1·73	2·14	3·31
23 Merry Wives of Windsor	54	9	41	3018	·46	1·08	1·67
24 Troilus and Cressida	463	133	439	3496	1·64	2·07	2·92
25 All's Well that Ends Well	349	138	316	2966	1·72	2·14	2·79
26 Measure for Measure	377	148	398	2820	1·77	2·06	2·92
27 Othello	679	268	694	3316	2·00	2·18	3·27
28 King Lear	580	243	691	3328	1·92	2·04	3·48
29 Macbeth	420	246	494	2106	2·11	2·14	3·09
30 Antony and Cleopatra	666	470	935	3059	2·24	2·00	4·09
31 Coriolanus	710	394	749	3406	2·16	2·30	3·52
32 Timon of Athens	334	145	404	2374	1·82	1·97	2·92
33 Cymbeline	799	393	1027	3339	2·15	1·93	3·98
34 Winter's Tale	675	330	699	3074	2·13	2·25	3·33
35 The Tempest	472	227	481	2062	2·10	2·18	2·88
36 Henry VIII	374	179	437	1167	2·12	2·00	2·69

x = redundant final syllables
y = full split lines
z = unsplit lines with pauses
w = total number of speech lines.

The chronological problem is essentially to divide the plays roughly into three or four groups and then to determine the succession of plays within the groups by means of external or metrical evidence. I have taken only thirty-six plays and divided them into three groups of twelve plays each by a casual examination of the metrical data except *Merry Wives of Windsor* which is alloted to the second group on the basis of external evidence. We now ask, what linear combination of these four variables x, y, z and w will make the differences between the groups as large as possible ? The problem, presented thus, easily lends itself to the application of the method in section 49·2 of Fisher's *Statistical Methods for Research Workers*. For the application of this method it is necessary that the variables should be normally distributed within the groups. It was found by the use of g_1, g_2 tests that x, z and w were normal (or very nearly normal with a slight kurtosis) within the groups as well as for the totals, while a logarithmic transformation was found necessary to make the distribution of y normal.

In order to find out the discriminant coefficients a, b, c and d appropriate to the four variables x, $\log y$, z and w, we first obtain the sums of squares for the groups and for the total. It is convenient to write the quadratic expression as a symmetrical 4×4 matrix. The whole has been multiplied by 36 to avoid too many fractions.

TABLE 2. SUM OF SQUARES FOR THE GROUPS.

	a	b	c	d
a	23405528	72006 ·544	28145110	–39356664
b	72006 ·544	245 ·353082	86360 ·672	20180 ·408
c	28145110	86260 ·672	33846614	–5040430
d	–3935654	20189 ·408	–5040430	44441222

TABLE 2A. TOTAL SUM OF SQUARES.

	a	b	c	d
a	53304443	106293 ·748	54129820	59116114
b	106293 ·748	351 ·193592	129605 ·852	80952 ·908
c	54129820	129605 ·852	68737592	55944284
d	59116114	80952 ·908	55944284	455405300

In order to find the values of the coefficients *a*,*b*,*c* and *d*, which will make the ratio of the first of these expressions to the second as large as possible, it is necessary to solve an equation of the fourth degree. If from each element of the first matrix, a multiple of the corresponding elements of the second matrix is subtracted, then the determinant so formed when equated to zero gives the required equation. The largest root of this equation can be most conveniently found by evaluating the determinant for chosen values of θ and then applying the method of divided differences. The following two tables show the values obtained at six chosen values of θ divided by 10^{20}, and the successive approximations to the largest root :

TABLE 3. VALUES OF Δ AT SIX CHOSEN VALUES OF θ AND THEIR DIVIDED DIFFERENCES

θ	determinant	divided differences			
		first	second	third	fourth
·0	·0000	5616 ·056	–30702 ·80	28897 ·0	293091
·2	1123 ·2112	–6665 ·062	–13364 ·61	263369 ·8	293091
·4	–209 ·8013	–12010 ·908	144657 ·29	497842 ·7	
·6	–2611 ·9829	45852 ·008	443362 ·94		
·8	6558 ·4188	223197 ·183			
1 ·0	51197 ·8555				

TABLE4. SUCCESSIVE APPROXIMATIONS TO THE LARGEST ROOT, FOURTH DIFFERENCE BEING293091

θ	determinant	first	second	third
1 ·0	51197 ·8555	223197 ·183	443362 ·94	497842 ·7
·7	–703 ·1012	173003 ·189	501939 ·94	585770 ·0
·71	–267 ·3945	43570 ·667	446319 ·04	618010 ·0
·715	–28 ·7447	47729 ·962	277286 ·31	593097 ·3
·7155	–4 ·0955	49298 ·490	285186 ·83	509712 ·9
715582	– ·0393	49466 ·139	288057 ·57	514279 ·8
·7155828	– ·0003	49490 ·015	288358 ·25	515916 ·1

To obtain the coefficients, the matrix for the total sum of squares is multiplied by this value of θ and subtracted from that fcr the groups, to give the following equations :

$$\begin{array}{rrrl}
-14738214.60000\ a & -4055.4338200\ b & -10589258.20000\ c & =46238128.4000\ d \\
-\ \ 4055.43382\ a & -\ \ 5.9550091\ b & -\ \ 6383.04647\ c & =\ \ 37739.1006\ d \\
-10589258.20000\ a & -6383.0164700\ b & -15340824.50000\ c & =45073197.4000\ d
\end{array}$$

As only three of the equations will be independent, if we take the value of *b* arbitrarily as unity, then the values of the remaining coefficients will be

$$a = \cdot00038648,\quad c = -\cdot00019325,\quad d = -\cdot00016664.$$

The discriminating score evaluated for each play is entered in column 5 of Table I.

It is rather disconcerting to find that we cannot proceed further with the problem without limiting the scope of its inquiry. There is no means of knowing the distribution function of the discriminating score, which would enable us to determine the succession of plays within the groups. In fact, it is not even possible to guage the power of our rough grouping, as the number of groups is too small to judge the linearity or otherwise of the group means of the score. It is, therefore, necessary to take only those plays whose dates are known or whose chronological order has been fixed at least with some certainty by reference to external evidence and to inquire if the changes in their discriminating score exhibit any linearity of regression. If it does, then it would be possible to determine the order of the plays of doubtful date or regarding the date of which there are differences of opinion among Shakespearean critics, by locating their position on the regression line. The order of the plays so taken is that given by Professor Chambers in his *William Shakespeare*, and I do not think that there would be any great measure of disagreement regarding this among Shakespearean scholars. Plays which are supposed to have been written during the course of the same year are bracketed and the averages of their discriminant score taken. The calculation of fitting Tchebysheff polynomials can be greatly simplified by the use of Table XXIII of Fisher and Yates Tables and are set out below in a compact form :

TABLE 5. CALCULATED VALUES BY FITTING POLYNOMIAL REGRESSION TO DISCRIMINANT SCORES

1	Hen VI 2 & 3	·58	11	Ham	·73
2	Hen VI 1	·81	12	Troilus	1·64
3	Rich III	1·38	13	AWW	1·72
4	Titus	·83	14	MM and Oth	1·89
5	RJ & TGV	1·32	15	Lear	1·92
6	Rich II & MND	1·10	16	AC	2·24
7	John & MV	1·44	17	Corio	2·16
8	Hen IV 1 & 2	1·10	18	Cymb.	2·15
9	MAAN & Hen V	1·06	19	WT	2·13
10	JC, AYL & TN	1·37	20	Temp.	2·10
			21	Hen VII	2·12

difference	sum
	1·73
·27	3·01
·66	2·78
·79	2·99
·48	3·36
1·14	3·34
·84	3·48
1·32	2·98
·75	3·51
1·29	2·91
1·54	2·70
	32·79

term	sums of products S(ξ'y)	regression coefficient	sum of squares
1	58·87	·076454545	4·500879064
2	−67·05	− ·000332105	·022267638
3	−145·78	− ·000336962	·049122364
4	−961·42	− ·000168071	·161586555
5	3311·21	·000027211	·090100913
S(u)=	32·79		4·823956534
ū =	1·5614		
S(u²)=	56·4687		
S(uū)=	-51·1992		residual mean square = ·4455
	5·2695		

The first term accounts for a substantial part of the total variation and the mean squares of the remaining four terms are all below the residual mean square. Thus a straight line adequately describes the changes and its equation is

$$u = \cdot7204 + \cdot07645v$$

It will be seen that *The Comedy of Errors, Love's Labour's lost* and *Merry Wives of Windsor* are early plays, while the position of *The Taming of the Shrew*, *Timon of Athens* and *Macbeth* is 5th, 14th and 18th respectively in Table 5.

The same method was tried with different groupings of the thirty six plays ; first they were divided into two groups consisting of the first sixteen plays mentioned by Meres and the remaining twenty ; and next they were classified as historical plays, comedies and tragedies including tragi-comedies. The appropriate discriminating scores for these classifications are set out in columns 6 and 7 of Table 1. For the former grouping, applying the method of Table XXIII of Fisher and Yates Tables, it was found that the first two terms accounted for about 71 per cent of the total variation, giving the equation of the regressional parabola as $u=1{\cdot}0118+{\cdot}1143\,r-{\cdot}0029\,r^2$. This also gave very nearly the same positions for the plays. The latter classification had to be abandoned totally as the variation due to linear regression hardly accounted for 55 per cent of the total variation.

There is no difficulty about placing the *Comedy of Errors* and *The Taming of the Shrew*, which are not mentioned by Meres and which, therefore, must have been off the stage at that time. The former seems to have closely followed the Henry VI plays. Its date is most probably the winter of the year 1591-92, which fits in with the allusion in III ii 125, first pointed out by Theobald, to the Civil war then raging in France . It is just possible that the play was written and produced shortly after the expedition of Norris and Essex, which was sent out by Queen Elizabeth in aid of the Protestant cause. As to *The Taming of the Shrew*, it seems to have been written shortly after *Titus Andrónicus*. Professor Chambers has come to the same conclusion, after identifying it, by the method of exclusions, with the *Love's Labour's Won* mentioned by Meres.

Furness gives in the Variorum edition a useful recapitulation of the dates arrived at by critics for *Love's Labour's Lost*. White, Fleay and Furnivall put it before 1590, relying on the want of experience and immaturity of style in the play, while Tieck, Sarrazin and Chambers give 1593-94 as the most probable date. The metrical analysis would class it with *the Comedy of Errors*. I would, therefore, place it in the spring of 1591-92, as the play has many parallels with Puttenham's *Arte of English Poesie* (1589) and Sir Philip Sydney's *Arcadia* (1590) and since it must have taken some time for Shakespeare to have come across these publications. There appears to have been some sort of remodelling or "augmentation" of the play by about 1593-94, as the parallels emanating from the Nashe-Harvey war would show.

As regards *The Merry Wives of Windsor*, its metrical composition shows unmistakable signs of earliness, unless its original version was by another author, as Wilson and Robertson would have us believe. In fact Fleay thought that it was a later and revised version of the *Jealous Comedy*, produced by the Alleyn company for Henslowe in January 1593. But jealousy is sufficiently a prominent motive in many other plays of Shakespeare and there is not much evidence to identify it exclusively with *The Merry Wives*. Fleay's view seems to fit in well with the most adroit and, therefore, unconvincing theory of Pollard and Wilson about the existence of a group of non-Shakespearean plays before 1593, which were revised and abridged hastily for the provincial tour of "Strange's" men which began in May 1593. Possibly the play was written early but 'put in cold storage' until it was required for production at a later date.

Dr. Bradley has placed *Timon of Athens* immediately after *King Lear* partly on account of their "similarity of temper and partly on metrical grounds". Professor Chambers puts it after *Coriolanus* as "it would be hard to find room for it in the already full year 1605-6". The statistical analysis gives the same year as the most probable time for its composition, and it may have preceded *King Lear*.

Regarding Macbeth, the metrical evidence seems to confirm the opinion of the Clarenedon editors that in all probability it was a new play when Forman saw it at the Globe Theatre in April 1610. Otherwise, they argue, he (Forman) would not have been in pains to make an elaborate summary of its plot. The allusions in Act II, Scene III to (1) the exceptionally low price of wheat in 1606 (Here's a farmer that hanged himself in the expectation of plenty, 1 ·4), (2) the doctrine of equivocation and the trial of the Jesuit Garnet in the spring of 1606 for complicity in the Gunpowder plot and (3) the fashion of wearing short hoses (Here's an English tailor come hither for stealing out of a French hose, 1.14) would merely fix a lower limit to the date of its composition. It does not seem right to suppose that William Warner added an account of the Historie of Macbeth to the new edition of his *Albions England* which appeared late in 1606, only after seeing *Macbeth*. Even more inconclusive are the echoes and parallels in the *Puritaine* or the *Widdow of the Watling Street* (1607) or in the *Tragedie of Caesar and Pompey* (1607), as echoes and parallels can work both ways and may owe their origin to a different source.

I thank Professor D. D. Kosambi of Poona for the valuable suggestions he made from time to time and for general guidance.

REFERENCES

1. E. K. Chambers : William Shakespeare
2. Dr. A. C. Bradley : Shakespearean Tragedy
3. R. A. Fisher : Statistical Methods for Research Workers
4. Arden and Variorum editions.

Paper received : 18 *August*, 1945

SAMPLE SURVEYS OF CROP YIELDS IN INDIA

By P. C. MAHALANOBIS

Statistical Laboratory, Calcutta.

HUBBACK'S EARLY WORK IN BIHAR AND ORISSA.

An account of early crop-cutting work in India carried out in 1923, 1924 and 1925, was given in an important paper on *Sampling for Rice Yield in Bihar and Orissa* by J.A. (later Sir John) Hubback of the Indian Civil Service (now retired) which was received for publication on 10 November 1926 and was published in 1927 by the Government of India as Bulletin No. 166 of the Agricultural Research Institute, Pusa. The method developed by Hubback was used a little later in Central Provinces by C. D. (now Sir Chintaman) Deshmukh, and Mr. P.S.Rau, two other officers of the Indian Civil Service. Hubback's paper also influenced work outside India. Prof. R.A.Fisher in his memorandum dated 2 March 1945 addressed to the Imperial Council of Agricultural Research, Government of India, stated :

> "The use of the method of random sampling is theoretically sound. I may mention that its practicability, convenience and economy was demonstrated by an extensive series of crop-cutting experiments on paddy carried out by Hubback (later Sir John Hubback, Governor of Orissa) more than 20 years ago over a greater part of the rice tracts in Bihar and Orissa. So far as I know these were the earliest crop-cutting experiments based on the principle of random sampling anywhere in the world. They influenced greatly the development of my methods at Rothamsted."

2. The bulletin has been, however, out of print for a long time. With the kind permission of the Government of India, we have much pleasure in reprinting it in the current issue of Sankhyā.

3. I had briefly referred to Hubback's paper in my memoir on Large Scale Sample Surveys (*Phil Trans.* Vol. 231 Series B). In the present note I propose to draw attention to a few salient points leaving it to the reader to read the original paper in full. Page references are everywhere to the present reprint.

4. Hubback's starting point was to criticise the unreliable character of official methods of estimating crop yields which were at that time and are still current in large parts of India. He observed :

> "All that is done at present is for the local police officers to make a guess, at which in succession the Subdivisional Officer, the District Officer, and the Director of Agriculture guess again". (p.282)

He pointed out that in the official method an attempt is made to select typical or representative fields for crop-cutting work which "depends entirely for its accuracy on the ability of the officer to select", and went on to say :

> "The method is comparable to estimating the average income of the population of a town by watching the streets for a few days and then picking out a man, who looked to be in average circumstances and discovering what his income is". (p.283)

Hubback also observed that in the official method "there is no possible way of estimating what is the probability that the result of such selections is within a given range from the true mean yield". He then stated :

> "The only way in which a satisfactory estimate can be formed is by as close an approximation to random sampling as the circumstances permit, since that not only gets rid of the personal element of the experimenter but also makes it possible to say what is the probability that the result of a given number of samples will be within a given range from the true mean". (p.283)

As far as I can trace this is the first explicit enunciation of the principle of random sampling for estimating the yield of crops.

5. Hubback refers to the earlier work in 1921 in which attempts were made to pick up fields at random, but points out that "in this method it was necessary to find out the date on which the crop on which field would be ready for harvest and to arrange to visit the village on that date." The practical difficulties involved in the above approach made Hubback search for another method in which, as he acutely observes, the distribution of sampling should be random in both time and space.

6. For crop-cutting work in India, cuts of the size of one-tenth of an acre have been standard for a long time. Hubback used a special apparatus giving a triangular cut comprising an area of only 1/3200 of an acre. His chief aim in reducing the size of the cut was to enable the investigators to collect the samples without much trouble or fatigue. In fact he noted that it is usually not possible to collect a large number of cuts of the size of one-tenth of an acre.

7. The margin of error of the final result received the closest scrutiny. From a careful analysis of the extensive data Hubback came to the conclusion that cuts of a small size were fully adequate. He observed :

> "It is no advantage to take a large number of samples from places very close together, where the crops will naturally be very much the same on the same day. The degree of accuracy is not seriously improved by such practice. This explains why there is no need to take large samples instead of the handy samples obtained by my method. A sample of one-tenth of an acre is merely 320 of my samples taken in juxtaposition .... Technically speaking there is very high correlation between the individuals of such groups of samples which makes the ordinary rule, that the standard deviation divided by the square root of the number of samples, quite inapplicable". (p.286)

This gives in a few sentences the central idea of the 'variance function' as used in our work nearly 10 or 12 years later.

8. Hubback was not satisfied in carrying out experiments in one or two small localities but covered a wide region during the three years 1923, 1924 and 1925. He stated :

> "The system has thus been tried in the Orissa deltaic tract, on the Chota Nagpur plateau of which Ranchi forms a part, in the broken country of Santal Parganas and Manbhum, and in the alluvial Gangetic plain of North Santal Parganas and Gaya". (p.287)

Hubback observed :

> "The method is applicable to small or large tracts. The same number of centres will probably give a slightly greater degree of accuracy for a small and a slightly less degree for a larger tract, because the standard deviation on the whole increases with the size of the tract, though not at all rapidly. (p.286)

9. It is remarkable that throughout his work Hubback had kept the question of cost prominently in view. With the help of actual experimental observations, he showed in fact that sampling by his method was entirely practicable and would give, at a comparatively small cost, final results with sufficient precision for most practical purposes.

10. Besides crop-cutting experiments, Hubback also made a number of important observations regarding estimates of crop acreage. For example, he emphasized that the harvested area was "evidently not the area planted or sown, since in practically all years some fields bear no crop worth cutting"; and noted that the difference "between the planted (or sown) area and the harvested area in a bad year, when the rains stop early in September, may well be 50 per cent all over a sub-division". He pointed out that the planted or sown area itself would vary from year to year. He actually observed :

> "It is usually believed that the estimate of cropped area is sufficiently accurate. It may be, but the belief is intuitive rather than rational". (p.287)

and pointed out

> "It would certainly be a great advantage if the harvested area could be obtained by some form of sampling". (p.287)

11. He himself tried a method in 1925 in the Santal Parganas which as far as I know is the earliest attempt to estimate crop acreage by the method of random sampling. He was not satisfied with the method, and he noted that "the field work is strenuous and tedious, and it is probable that it would be in practice shirked, and results fudged". He mentioned other difficulties, but concluded with the observation :"But it is still possible that some method on similar lines may prove practicable". It is interesting to note that he had also pointed out the importance of taking into consideration the area under field ridges (*ails*), and of making reliable estimates for the loss of weight at the stage of husking paddy.

12. In an appendix he gave a detailed investigation into the effect of deviations from "simple" sampling. As far as I am aware, this is the earliest discussion of what would be now called the method of determining the error when the sampling is done in more than one stage. Hubback had also investigated how far frequency distributions of yield rates conformed to the normal curve.

13. For a proper appreciation of the real significance of Hubback's work, it must be remembered that it was so late as in May 1924 that the Bureau of the International Institute of Statistics appointed a commission for the purpose of studying the application of the representative method of statistics, and that this report was presented in 1925 and published in 1926. In this report considerable prominence was given to the method of purposive selection, but the importance of method of random sampling had not been fully realised. As already noted, Fisher's own work at Rothamsted had been directly influenced by Hubback's paper. Hubback's work was also the starting point of much of the work done by the Indian Statistical Institute in Bengal and Bihar since 1938.

14. Against the historical background sketched above, Hubback's work must be considered to be a most remarkable achievement. The paper itself is quite small and gives

in a short compass a masterly discussion of the subject. It truly constitutes a classical paper which deserves careful study even after the lapse of 20 years.

DESHMUKH'S WORK IN CENTRAL PROVINCES.

15. Immediately after the publication of Hubback's paper in 1927 similar work was taken up in 1928 in Raipur district of the Central Provinces of India by C.D. (now Sir Chintaman) Deshmukh, Governor of the Reserve Bank of India and at present President of the Indian Statistical Institute, and was continued in 1929 and 1930 . A brief account of this work with special reference to applications in the assessment of land revenue was given by Deshmukh in his report on the Settlement of the District of Raipur, 1926-1931 published by the Government of Central Provinces. A little later Hubback's method had been used by Mr. P.S.Rau, I.C.S., for the sampling of rice yield in Bilaspur.

16. At the time of my visit to Nagpur during the annual session of the Indian Science Congress in January 1945 I happened to be staying with Mr. Rau. At my request he very kindly traced the original papers and sent me 21 books containing original entries out of which 19 books relate to Deshmukh's work in Raipur, and 2 books to Rau's work in Bilaspur. After a good deal of search in Government offices I was informed that the bulk of the material containing the observations made in Bilaspur district had been unfortunately destroyed some considerable time ago.

17. Deshmukh's material consists of 1517 cuts in 1928, 1447 in 1929, and 1920 in 1930 making up a total of 4884 crop cutting experiments altogether. (I may note in passing that the figure 4384 for the total number of cuts given in paragraph 93 on p.28 of the Raipur Settlement Report is evidently a printing mistake which had almost certainly arisen from the substitution of the figure 3 for the correct figure 8.) In the original record, entries were available relating to the name of the village, the variety of paddy, irrigation factors etc. The material is being analysed in the Statistical Laboratory and a separate note will be published later. The village entries make it possible to carry out a two-stage analysis. At my request Mr. C. R. Rao undertook this work, and has supplied the following material.

18. Let σ_1^2 be the variance between the average rates of yield per sample-cut of villages, and σ_2^2 the variance of the rates of yield per cut within a village, and let n_1 be the number of villages included in the survey from each of which n_2 fields are sampled. The estimated variance of the average rate of yield per cut for the whole region is given by

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2}$$

The following table gives the values of σ_1 and σ_2 and the mean yield in tolas per cut based on the material collected by Deshmukh in Raipur in 1928, 1929, and 1930.

Table 1. Estimated Values of Mean Yields and Standard Deviations : Raipur 1928-1929.

year	number of cuts	estimated values of		mean yield per cut
		σ_1	σ_2	
1928	1517	5·24	6·06	12·82
1929	1447	2·73	7·34	17·88
1930	1920	4·96	6·36	14·06

19. From the above data it is possible to determine, for any given number of cuts per village, the number of villages which would have to be sampled to reduce the percentage

error of the estimated mean yield to any assigned value. The following Table 2 gives relevant figures for a percentage error of 3%.

TABLE 2. ESTIMATED NUMBER OF VILLAGES AND SAMPLE-CUTS
(PERCENTAGE ERROR=3%)

year	number of villages requiring to be sampled for				
	1 cut	2 cuts	3 cuts	4 cuts	5 cuts
1928	434	314	269	248	236
1929	214	120	89	73	64
1930	366	253	214	196	184

20. The important point to be noticed is that there is a substantial reduction in the number of villages to be sampled when the number of sample-cuts is increased from 1 to 2 or 3. With small cuts of the size used by Hubback, Deshmukh and Rau (or by the Indian Statistical Institute more recently) there is no difficulty in collecting a number of cuts in the same village or in a group of adjoining villages. The use of sample-cuts of a small size is thus distinctly economical.

21. Another point is also worth stressing. From Hubback's own data and also from the calculated values of the variances based on Deshmukh's material it is clear that only a moderate number of sample-cuts would be quite sufficient to give reliable estimates of the mean values of the rate of yield per acre for a whole province. For example, using Deshmukh's data for 1930 (the year in which variances were highest) it appears that less than 100 villages with one or two cuts in each distributed suitably over the 9 different districts in the Central Provinces would give the provincial mean rate of yield with an error less than 2 per cent which should be adequate for all practical purposes, and would in any case reach the limit of physical fluctuations.

LATER DEVELOPMENTS

22. In 1938 extensive crop-cutting work was done in Bengal (under the general supervision of Mr. H. P. V. Townend, then Development Commissioner, with the technical help of the Statistical Institute) with a square wooden frame including an area of 27.04 sq. ft. or just about double the size of the triangular frame used by Hubback. Since then a good deal of intensive studies of crop-cutting work have been made by the Indian Statistical Institute. The chief objects of these investigations have been four-fold :

(1) to settle the standard size of the sample-cut and appropriate specifications for actual harvesting in the field ;

(2) to settle the number and distribution of sample-cuts over the region under survey in order to attain any desired degree of precision in the final results ;

(3) to study the most efficient organization of the human agency including the control of recording mistakes ;

(4) to study possibilities of using auxiliary measurements having correlation with the final yield of dry grain in the case of cereals and dry fibre in the case of jute.

23. *Size and shape of sample-units.* During the last 5 or 6 years extensive experiments have been made on sample cuts of various shapes and sizes, and using many different procedures among which may be mentioned the following : Hubback's triangular frame of size 13.6 sq. ft. ; Townend's square frame of size 27.04 sq. ft. ; cuts of square and rectangular

shapes demarcated by ropes and pegs and of various sizes ranging from 1 sq. ft. to about 1/9th of an acre ; cuts of circular shape with radius of different lengths such as 2′, 4′, 5′-7½″, 8′ etc. harvested with the help of an instrument first designed by Mr. J. M. Sen Gupta of the Statistical Laboratory and then improved in various ways. The circular cut is being used at present as a standard procedure in Institute work.

24. An important question is whether the use of cuts of a small size is fully representative and unbiased. This has been and is being studied by the Institute practically every year since 1939. From the theoretical point of view, cuts of a large size would of course lead to a relative under-sampling of the portion of the field near the boundaries. The undersampling arising in this way varies roughly as the ratio of the size of the cut to the size of the whole plot assumed to be of a square shape. This is a sampling bias, and is to be carefully distinguished from the bias, if any, which may arise from the personal equation of the investigator in the actual process of harvesting the cut which is the point under consideration here.

25. Broadly speaking, it was found by the Institute that with cuts of a very small size, when these are demarcated by pegs and ropes, there is a tendency for over-estimation of the rate of yield per acre. It was found however that this tendency decreases as the size of the cut is increased, and practically vanishes beyond about 50 sq. feet under normal conditions. It is possible that when a small cut is demarcated by pegs and ropes there is tendency for the investigator to be over careful and to pull in a few of the bordering plants which should properly be allowed to fall outside the rope.

26. With circular cuts of a very small size there is some evidence of a small bias sometimes creeping in which may however be both positive (over-estimation) or negative (under-estimation) depending on the investigator. On the whole, when the work is done by a large number of different investigators, these tend to cancel out so that results obtained with the circular cut, as far as can be judged from evidence at present available, appear to be without bias. Another advantage of using cuts of a circular shape is that no bias can arise due to varying directions along which the crop is sown.

27. *Variance function.* The variance function, that is, the decrease in variance with increase in the size of cuts, has been studied in great detail in the Institute. As pointed out by Hubback, there is comparatively little gain in precision by increasing the size of the cut owing to the existence of positive and usually high correlation in the yield of plants in adjoining portions of the same field. This has been fully borne out by extensive work done by the Institute in recent years. From the point of view of precision, nothing is to be gained, therefore, by using cuts of a large size. As pointed out by Hubback, cuts of one-tenth of an acre usually used in Government Agricultural Departments, are entirely unnecessary.

28. A far larger number of cuts of a small size can be harvested in the same time. Deshmukh's work in 1928-29-30 had shown (and has been corroborated by subsequent work by the Statistical Institute) that 4 or 5 sample-cuts collected in the same village or in a group of adjoining villages would give results of about the same precision as single cuts collected from about twice the number or more of villages. As the greater part of the total cost arises from journeys which have to be undertaken from one village to another, it is clear that a reduction in the number of villages (made possible by using cuts of a small size) would mean much less expense when the work is done by a field staff moving from village to village. If the sample-cuts are harvested by stationary investigators (who do the work in the near neighbourhood

of their normal place of residence) even then each such investigator would be able to collect in the same time a much larger number of sample-cuts of a smaller size which would lead to a distinct gain in precision.

29. Available evidence thus distinctly indicates that sample-cuts of a small size would be more economical. Another important consideration is the greater ease and convenience with which such cuts can be harvested, especially with the circular implement, as this is bound to improve the quality of the field work.

30. *Density and total number of sample-units.* As already noted, since Hubback's time experimental evidence has continuously accumulated showing that coefficients of variation for cuts of a small size are quite small and of the order of only 50 per cent. Two to three thousand sample-cuts properly randomized in both space and time should thus, in theory, supply a mean value of the yield per acre with a precentage error of the order of 1 per cent which would be amply sufficient for all practical purposes. The use of cuts of a comparatively large size is thus entirely unnecessary so far as precision is concerned.

31. A second point had also emerged from Hubback's early work, namely, that variances or coefficients of variation for large tracts are only slightly greater than the variance or co-efficients of variation for much smaller tracts into which the whole region may be sub-divided. This has been also borne out by Deshmukh's data, and by other material collected in the course of more recent surveys. The same number of sample-cuts distributed at random over a single district, or over the whole province would thus supply mean values with percentage errors which would not be very different. The cost of field operations would thus depend almost directly on the number of districts or regions for which separate estimates have to be obtained.

32. *Recording mistakes.* As already noted, considerations of cost make it necessary to use multi-stage sampling in crop-cutting work. Units at the higher stages, namely, zones, cells, villages or fields can be and are selected at random in advance in the Statistical Laboratory. The location of the sample-cut itself within the field in the final stage must however be done by the investigator ; the sample-cut has also to be harvested, threshed and weighed by him. In the survey of crop acreage, it is possible to have the same field inspected by two or more independent investigators. The same sample-cut can never be harvested more than once. Possibilities of the occurence of unconscious bias as well as of gross negligence or deliberate dishonesty are therefore much greater in crop-cutting work.

33. This indeed has been found to be the greatest obstacle, under working conditions in Bengal and Bihar, to the securing of reliable estimates of rates of yield of crops per acre and the total outturn of crops. This problem has not yet been solved, but is being continually studied, and attempts are being constantly made to secure unbiased primary records in crop-cutting work. The method of independent net-works of samples is of course of great value in supplying an over-all control. Various other methods are also being tried and are being used with different degrees of success. A brief description of some of these methods is given below :

34. *Interpenetrating net-works of samples** (*replication*).—In crop survey work the Indian Statistical Institute has been using for the last 6 years replication in the form of two independent net-works of sample-units. In each zone or region (within which the act of random location of sample-units is separately completed) the sample-units, as they are located, are numbered consecutively. It is thus possible to divide the whole set into two (or more)

---

*This phrase is due to Prof. R. A. Fisher.

independent but random sub-sets. For example, the odd numbered sample-units would form a random set as also the even numbered sample-units. These two sets would be interpenetrating and would each cover the whole zone or region. Informaton for each sub-set is collected by a different party of field investigators who work under different inspecting staff. When two such interpenetrating net-works of sample-units are used, it is possible to obtain two independent estimates (of crop acreage or of crop yield) for each region. The difference between these two independent estimates immediately supplies a good idea of the effective margin of error and thus indicates to what extent the survey had been carried out under statistical control.

35. *Control through variance function.* In crop-cutting work (for estimation of crop yields per acre) another kind of control has been found most useful. In the Institute method the standard procedure is at present to obtain cuts of a circular shape of different sizes harvested by a compact instrument which implement consists essentially of a folding arm which can be opened out to various lengths (2′, 4′, 5′-$7\frac{1}{2}$″, 6′, and 8′). A pivot with a stump about 12″ long is driven vertically into the ground at a point selected at random, and the arm is placed on a socket arrangement on the stump in such a way that it can swing freely in a horizontal plane (i.e. in a direction perpendicular to the pivot) and describe circles of various sizes. A small pointer is attached at right angles at the free end of the arm, and as the arm is rotated, all plants falling inside the pointer are harvested. The cut of the smallest size is naturally first collected, and the grain threshed and weighed. The arm is then drawn out to the next larger radius and the additional crop in the annular strip is harvested, threshed, and weighed. The same process is repeated for circles of larger size. In this way the weight of grain for 3 or 4 different sizes of cuts (each of circular shape and with centres at the same point on the ground) are obtained.

36. A comparison of mean values based on different sizes of cuts shows whether or not any bias had arisen in using different sizes of cuts in the process of harvesting. Cuts of different size would, however, in general show decreasing variance with increase in size. A comparison of variances for cuts of different sizes thus supplies another valuable control. For example, it is sometimes found that mean values based on cuts of different sizes are in excellent agreement (often much better than one could expect from probability considerations), and variances of yield rates for cuts of different sizes are also about equal. In such cases it is practically certain that cuts of only one size had been actually harvested and records for cuts of other sizes were obtained by paper calculations. Multiple cuts of other shapes and sizes have also been used in crop-cutting work, and found to be quite useful.

37. In crop-cutting work it is practically always necessary to use a multi-stage design as already explained. The selection at random of units at the higher stages, namely, zones, cells, villages or fields is usually done beforehand in the Statistical Laboratory. The selection of the sample-cut itself, at the final stage, must however be done by the investigator on the spot. For this purpose tables of random numbers are supplied and the investigator is required to move an appropriate number of steps along and across the field according to the random numbers picked up in serial order from the table. Sometimes the investigator is asked to turn to the right or to the left or to get a particular segment out of a number into which the field is divided, depending on whether the toss of a coin shows head or tail. Each investigator, in such cases is required to keep a record of successive throws of heads and tails ; and as fairly long series are usually obtained for each investigator, it is easy to test whether the heads and

tails form a random series or not. If, as is sometimes found to be the case, the series is significantly non-random it is reasonable to think that the field work was not done properly. It is true, of course, that coin tossing may have been done properly but not the field work. Still it is often useful to have even such partial checks.

38. *Field organization.* In crop-cutting work the great snag is that harvesting starts at different times in different regions and even in different villages within the same region. Once started, the harvesting is often over very quickly, sometimes in the course of a fortnight or so. Another difficulty, which is peculiar to paddy, is that different types of paddy are sometimes grown in the same region or in the same village, so that harvesting of paddy of one type or another may continue in the same village from June or July till December or January. In fact in Bengal, harvesting of paddy goes on practically in every month of the year in one part or another of the province.

39. A well organized staff of *patwaries* or village revenue officers exist in the temporarily settled provinces in India but no such staff is available in the permanently settled provinces like Bengal and Bihar. In the Bengal Crop Survey the crop-cutting work is therefore done by the investigators engaged in the acreage survey. The total number available being only 300, the investigators have to move from one village to another in accordance with a fixed programme. Owing to large fluctuations in the date of harvesting, it is inevitable that the investigators should sometimes find that the crop has already been harvested before they reached a particular village or that the crop would still require a number of days to be ready for harvesting at the time of their arrival in any particular village or locality. In this situation the standard programme settled beforehand has to be modified in many instances. In theory, this might take away somewhat from the ideally perfect randomization of the sample. In practice, this appreciably reduces the number of sample-cuts which can be actually collected at the proper time.

40. After struggling with the problem for many years it is becoming clear that crop-cutting work to be done properly must be carried out by a comparatively large number of investigators who would watch the crop as it grows, and collect the sample-cuts at the right time from fields situated in the neighbourhood of their normal places of residence. The total volume of work required from each investigator would not be large, and the work can be easily done as a part-time occupation. With the circular cuts of the size used by the Institute it is possible to harvest up to 5 sample-cuts (each consisting of a set of 3 or 4 multiple-cuts of different sizes) in the course of a day. One investigator working in a group of 6 or 8 adjoining villages can easily collect 20 or 30 sample-cuts in the course of a total working time of 6 or 7 days. If a staff of two or three thousand part-time workers were available scattered in a suitable manner over the province it would be quite sufficient if each of them did crop-cutting work for about a week in each crop season. This would supply mean rates of yield per acre for individual districts with a percentage error of the order of 1 or 2 per cent.

41. There are only two ways in which the help of a sufficiently large number of investigators can be secured for crop-cutting work at the proper time. One and probably the best plan would be to include crop-cutting work as an integral part of a comprehensive scheme of multi-purpose sample survey set up throughout the province. In such a scheme a nucleus of whole-time workers would be maintained and spread over the province. A large number of workers would also be employed who would do crop-cutting as well as other sample survey work as a part-time occupation. Employment for crop-cutting work alone would not be

economical, but there would be no difficulty in offering sufficiently attractive remuneration for part-time work to the extent of say 2 or 3 months in the year.

42. The only other alternative is to employ some Government agency already existing in the province such as the Jute Regulation staff in Bengal or the subordinate postal staff. In the temporarily settled provinces there is also the possibility of using patwaries. These are important organizational problems which require to be studied on scientific lines without further delay.

43. *Driage.* The possibility of using auxiliary physical measurements (which have appreciable correlation with the yield of dry grain) has been and is being investigated. For example, in the case of paddy and wheat it has been found that there exists a high and quite steady correlation between the weight of crop immediately after harvesting (called "green" weight) and the weight of crop after drying. There are of course, minor fluctuations from one season to another or from one region to another in the same season. Arrangements are therefore made, in about 10 per cent of the sample-cuts, to obtain the weight of crop after drying as well as the weight immediately after harvesting. On the basis of the material collected in this way it is possible to work out reliable methods for estimating the weight of dry grain from the much more extensive material for green weight.

44. *Mixed crops.* In certain parts of India, especially in the rabi (or winter) season a number of different crops such as barley, wheat, gram and/or other pulses are grown on the same field. The situation is further complicated by the fact that the relative proportion of different components of the mixed crops vary widely from field to field. Special methods had to be developed for estimating the total production of each component crop in such cases. Incidentally, it may be observed that relatively a much higher yield is obtained when a particular crop is sown in comparatively thin proportions. Full details have been given in a report on the crop survey in Bihar in the *rabi* season of 1943-44. in *Sankhya*, Vol. 7, Part 1, 1945.

45. *Husking.* In the case of paddy the weight of clean rice after husking is conventionally accepted to be two-thirds of the weight of dry paddy before husking. This ratio has however been found to fluctuate from region to region and from season to season. Arrangements are therefore made to measure directly the husking ratio at a number of places scattered over the province. Evidence is accumulating that the driage as well as the husking ratio depend on conditions of cultivation. These are being gradually studied.

46. *Jute.* In the case of jute, in the same way, it has been found that the weight of the jute plants measured immediately after harvesting has a high correlation with the weight of dry fibre extracted after retting and drying the plants. Here also, in crop-cutting work, the weight of plants in the sample-cut is recorded immediately after harvesting, and the extraction ratio is experimentally studied in the case of a small proportion of sample-cuts.

47. *Cinchona.* Possibilities of using physical measurements on living plants are also being explored in many cases. For example, in the case of cinchona plants it has been found that it is possible to forecast the total yield of bark with sufficient precision from physical measurements of girth of the plant at a height of six inches above ground, height of the plants and number of stems on living plants.

48. *I. C. A. R. Scheme.* I may at this stage make a brief reference to the crop-cutting work by the Imperial Council of Agricultural Research of India which was started on the wheat crop in the United Provinces in 1944, and has since then been extended to both paddy and wheat in other provinces. No general account of the work has been published, but I have

receiveb form the I. C. A. R. preliminary reports relating to two or three provinces. It is not possible therefore to make any detailed comments, but a few general observations would not be out of place.

49. The first point to be noted is the use of random samples. There is, of course, complete agreement between the I. C. A. R. and I. S. I. methods in this respect. It is a matter for congratulation that through the efforts of the I. C. A. R. the principle of random sampling is likely to be adopted in practice throughout India at an early date.

50. There are however certain important difference in the methods used by the Indian Statistical Institute and the I. C. A. R. which may be broadly grouped under four heads :

(1) *Size of the sample-unit:* In the I. S. I. method, the sample-cuts are comparatively small in area usually ranging between about 50 and 100 sq. feet. The I. C. A. R. sample-cuts are of a much larger size, of the order of 900 sq. feet or more.

(2) *Harvesting procedure.* I. S. I. sample-cuts are of circular shape obtained with the help of the special implement described earlier, while I. C. A. R. sample-cuts are of square or rectangular shape demarcated by pegs and ropes or chains.

(3) *Use of sub-samples as control.* In the I. S. I. method the crop is harvested in the form of three concentric sub-samples and the yield from each sub-sample is separately weighed and recorded. The variaton of yield between such sub-samples supplies an excellent internal control which is not available in the I. C. A. R. method.

(4) *Use of interpenetrating net-works of sample-units.* As already mentioned, in the I. S. I. method great importance is attached to arranging the field work in the form of two interpenetrating net-works of sample-units and collecting data for each such set of sample-units by entirely different parties of investigators. No such replication is available in the I. C. A. R. method.

51. The Institute view has been that the method developed by it has distinct advantages. It has good statistical controls at the point of collection of the data, and is also less expensive when the work is done through the same type of human agency. It is difficult however to make direct comparisons between relative efficiencies of the two methods for the reason that these have been used in different parts of India. The I. S. I. method has been used mostly in the two provinces Bengal and Bihar where the land revenue is permanently settled and where no revenue or other village staff under Government control is available for doing the crop-cutting work. The I. C. A. R. method on the other hand has been so far used in temporarily settled provinces like the Punjab or U. P. which all have a well organized staff of *patwaries* or village officials under Government control whose part-time services are available for doing the crop-cutting work.

52. There has been some controversy in India regarding the relative advantages and disadvantages of the I. S. I. and the I. C. A. R. methods for estimating crop yields. As I happen to be closely associated with the I. S. I. method , it wouid not be advisable to make any detailed comments until full reports of I. C. A. R. work become available. I may mention, however, that for some considerable time I have been pressing on I. C. A. R. authorities the need of carrying out crop-cutting work by both I. S. I. and I. C. A. R. methods in the same region with a view to studying the relative efficiencies of the two systems. I have submitted definite schemes for this purpose which are under consideration by the Government of India.

JOINT OR DIRECT ESTIMATION OF CROP PRODUCTION

53. Since 1938 I had been advocating the sample survey of crop acreage and crop-cutting work being taken up jointly, but the opportunity occurred only several years later to carry this into practice. A brief description of the method as actually used during the survey of paddy in 1945-46 in Bengal would explain the procedure. The number of grids in which crop-cutting work is done is necessarily much smaller than the number in which the acreage survey is carried out. In 1945-46, about 7000 grids were selected at random for crop-cutting (but in a suitably zoned form), and a second survey of the area actually under paddy at the time of harvesting was carried out. In this way the total acreage under paddy within each grid was known. Multiplying by the rate of yield of paddy per acre (as determined by crop-cutting work carried out in the same grid) it was possible to calculate the total production of paddy per acre for each individual grid. Taking the average within each cell, the total production for each cell could be then directly calculated, from which total production for larger tracts or for the province as a whole could be easily built up.

54. In 1945-46 in Bengal the total production estimated directly in this way was about 6.52 million tons of rice (not in husk). The corresponding estimate obtained by multiplying the acreage by the mean rate of yield was about 6.73 million tons. Incidentally, it may be mentioned, that the second survey at harvest time indicated an appreciable (about 6 per cent) decrease in the area harvested as compared to the area sown.

55. *Conclusion.* A good deal of work has been done and some progress made in India in developing an efficient sampling technique for the estimation of crop yields and total production of crops. A good deal, however, still remains to be done. The active co-operation of statisticians working in different parts of the country is essential for rapid advance.

Paper received : 20 March 1946.

SAMPLING FOR RICE YIELD IN BIHAR AND ORISSA

By J. A. HUBBACK, I.C.S.

The need for reasonably accurate statistics of the human population of a country has long been recognized throughout the civilized world. The idea that statistics of the production which supports that population are of well-nigh equal importance is rapidly gaining general recognition. It is, therefore, perhaps superfluous to argue in its favour but it may be useful to indicate what objects a well-founded estimate of the yield f of rice can fulfil.

In the first place, since rice occupies a little more than half of the cultivated area in the provinces of Bihar and Orissa, employs 4/5th of the total population for six months of the year and feeds the greater part of it throughout the year, its supreme importance in any survey of production can scarcely be denied. It probably provides at least one-half of the rural contribution to the "provincial dividend", and that contribution is still vastly more important than that of the urban and industrial workers. Hence if a survey of production is desirable, good statistics of rice yield are essential.

Apart from the uses to which those who direct or influence the policy of the State may put the statistics of production, there is no doubt that traders want information of crops as they come forward. Good information enables them to operate with more confidence and not only saves them from alternating between prosperity and financial disaster, but greatly benefits the public by stabilizing prices. For this purpose it is not enough to know what is the production of the whole province, but knowledge of the conditions of the districts or even smaller units is very valuable. Lastly, reasonably accurate knowledge of the real normal yields of important crops even for fairly large tracts would give guidance to courts who are required to estimate the value of produce for the decision of rent suits, for commuting produce rent or for settling fair cash rents. Hitherto lack of such knowledge has frequently led to considerable injustice.

The statistics of area under the rice crop are claimed to be fairly accurate. They are founded ultimately on the figures obtained from detailed measurement in a particular year on the ground. No doubt new land has been brought under rice in some places, the year of survey may have been a particularly favourable or unfavourable one, or to a limited extent other crops may have replaced rice. Allowance is made for these factors year by year, and perhaps there is no very large error. The question is examined further in the latter part of this bulletin. But in the figure for yield there is certainly room for bad mistakes. This has been fixed under the advice of the Director of Agriculture at a definite figure as normal, and is varied year by year by applying for each district a percentage, indicating how far the crop of the year falls below or rises above the normal. Now in the first place it is exceedingly probable that the normal rice yield differs materially between districts. But, apart from that, there is no machinery for fixing with any reasonable accuracy the percentage of the year.

EDITORIAL NOTE— This paper was originally published in 1927 as Bulletin No. 166 by the Agricultural Research Institute, Pusa, and is being reprinted by the kind permission of the Government of India. This is probably the earliest account of a scientific experiment to determine crop-yields by means of a sample survey.

All that is done at present is for the local police officers to make a guess, at which in succession the Sub-divisional Officer, the District Officer, and the Director of Agriculture guess again. When it is considered that the percentage depends ultimately on the effect of the weather on very various soils, cultivated with varying degrees of skill and enterprise, planted with different kinds of rice, protected by irrigation works of greater or less efficiency or completely unprotected, liable to or immune from crop pests, and finally harvested over a period of nearly three months, it becomes apparent that the guessing ability of the officers concerned has to be remarkable. Unfortunately, too, not one of them has the least chance of finding out whether his guess was fairly right or wildly wrong. Hence the existing statistics of rice production are, I believe, the result of applying to a fairly accurate figure of area, an arbitrary standard of normal yield, and a pure guess of the condition of the year. I am convinced that by an expenditure of a sum of money, that is not excessive in view of the importance of the matter, the last two and perhaps the first could be replaced by a scientific estimate, in which the possibilities of error can be gauged. Incidentally the scheme proposed, though it would not completely solve the problems of the courts to which reference has been made, would go a great way to do so.

I return to the point of the normal yield. This is based on a number of crop-cutting experiment made at different times in different places, and it is to the defects of these experiments, with the consequent uncertainty of the conclusions based on them, that I now wish to draw attention.

The commonest method of estimating yield, apart form pure guessing fom the appearance of the field at harvest time, is that described in the standard account of the danaband (appraisement) system of rent payment. It consists of the tenant cutting a small area where the crop is thinnest and the landlord where the crop is heaviest, mixing the two together and taking the result as a fair sample. To secure accurate results this presupposes that the true mean lies midway between the highest and lowest sample, a condition which is very rarely found in practice. The results so obtained are almost always far too high, i.e., unfair to the tenant. In an acre field of rice bearing a mean yield of 15 maunds it might well be possible to find an area of (1/20)th of an acre yielding 40 maunds an acre, but it would be obviously impossible to find another bearing minus 10 maunds an acre. The method is not far different from attempting to estimate the average income of a population by taking half the sum of the incomes of the poorest man and the richest man in it.

A second method is that at present prescribed. An officer of some standing visits the tract, of whose yield he is required to form an estimate, and selects a fairly large area, usually (1/10)th of an acre, as containing an average crop. This he cuts and carries and in, due course threshes and weights. If he has time he makes two or more experiments, but it is obvious that he cannot usually manage to make a large number of such experiments. There are two main objections to this method : the first is that it depends entirely for its accuracy on the ability of the officer to select. He has to make his selection from a large number of fields growing different varieties at different points of maturity. It is very easy to be misled as to yield by the strength of straw, heavy straw being by no means always correlated with heavy yield of grain. Again it is difficult to give proper weight to the fields which will give little or no yield at all, and even in a normal year such fields are by no means rare. These factors make accurate selection very difficult. The second objection is even more important, viz., that there is no possible way of estimating what is the probability that the result of such selections is within a given range from the true mean yield. The method is comparable to estimating the average income of the population of a town by watching the streets for a few days and then picking out a man, who looked to be in average circumstances and discovering what his income is.

The only way in which a satisfactory estimate can be formed is by as close an approximation to random sampling as the circumstances permit, since that not only gets rid of the personal element of the experimenter, but also makes it possible to say what is the probability that the result of a given number of samples will be within a given range from the true mean. To put this in definite language, it should be possible to find out how many samples will be required to secure that the odds are at least 20 to 1 on the mean of the samples being within one maund of the true mean.

This is the problem to be solved. Before describing the progress made towards a solution, it is desirable to explain some terms used in modern mathematical statistics, on which the solution depends. These will, no doubt, be quite familiar to many of my readers, to whom I apologize for the digression. The first term is the arithmetical "mean", which is the real object of our search. It is, of course, the sum of all the individual results of the cuts or samples made divided by the number of cuts or samples. The true "mean" is the figure which would be obtained if the number of samples were increased, so as to cover the whole subject matter of sampling exhaustively. The whole subject matter of sampling is known as the "population" viewed as made up of a very large number of samples with varying values. The manner in which the values of the samples are numerically arranged is called the "distribution". Clearly the distribution may in some populations be very much wider than in others. The number of samples having a given value, or having a value lying within given limits, is called the "frequency" for that value or for the range of those limits. The range of limits is known as the "class interval", and, for convenience of calculation, values of samples , when these vary continuously, are lumped together in equal class intervals and treated as if they all fell at the middle point of their class interval. A most important term is the "standard deviation" of the distribution. This is a measure of the amount of dispersion or "scatter" about a particular value usually about the mean. It is a perfectly definite value expressed in the same unit as that in which the value of each sample is expressed. It is obtained as follows. The value of each sample differs by a definite figure from the particular value, about which the dispersion is sought, e.g. from the mean. This difference is squared and (where the samples have been lumped together in class intervals) multiplied by the frequency of the class interval, The results are added together and divided by the sum of all the frequencies, and the square root of that result is the standard deviation. A simple illustration will explain the process. Suppose that 8 samples give the following results.

frequency	value	difference from the mean which is 6	square of the difference multiplied by the frequency
1	4	— 2	4
4	5	— 1	4
1	6	0	0
0	7	+ 1	0
0	8	+ 2	0
2	9	+ 3	18
total 8			total 26

The standard deviation is the square root of 26 divided by 8 or 1.8 nearly.

There remains to be explained yet one more term, "the normal curve of error". This is a graphic representation of a certain class of frequency distribution, which has, among other properties, that of symmetry about the mean. To this class many of the distributions actually found in practice more or less closely conform. It has the very great advantage that the chance that a particular sample will fall within a definite range on either side of the mean can be calculated at once from standard tables, and this is what makes it so widely used in statistical practice. The properties of other classes of frequency distributions have been investigated, but their use involves difficult and laborious calculation, which makes them unsuitable for many kinds of work. It is not infrequently assumed that the determination of the standard deviation of any frequency distribution at once enables one to say what are the odds that the mean of sampling falls within given limits on either side of the true mean. This is untrue, and it is always necessary to see that the distribution under scrutiny conforms closely to the normal curve of error, before making deductions, which are true for that distribution but not for others.

Evidently then before a practical method of obtaining a satisfactory value for the mean yield of the rice crop from a system of sampling could be recommended, it was necessary to get a clear idea of the distributions, which random sampling will give. The first attempt was made under the direction of Mr. Dobbs, Director of Agriculture, Bihar and Orissa, in 1921 in the Nawada thana of Gaya. His method was to divide up the map into equal squares, in order to pick out the villages in which samples were to be taken, and then to sample the fields bearing the survey number 50, or the nearest fields planted with rice. For this method it was necessary to find out the date on which the crop on each field would be ready for harvest and to arrange to visit the village on that date. The practical difficulties involved made me search for another method, more automatic in the distribution of sampling in time and space. It is unfortunately impossible to use the result of Mr. Dobbs' experiment, because three samples were taken in each field and the results lumped together. This has disguised the frequency distribution so much as to make deductions extremely hazardous.

In 1923 with the help of Mr. Davies, the Settlement Officer of the Santal Parganas district, 400 random samples were obtained in a part of Godda thana. The sampler spent from the 4th December to the 22nd December travelling from village to village through the area comprising about 100 square miles. Wherever he found, on the day of his visit, the crop being actually harvested, he took a sample. In the following year the same method was applied under the supervision of the Settlement Officers, Messrs. Davies, Mansfield and Gokhale, in the Godda thana of Santal Parganas, the Jaipur thana of Cuttack, and the Purulia thana of Manbhum. The only difference was that four samples instead of one were taken where practicable, on each field in which the harvest was in progress.

In 1925 I obtained the permission of the Bihar and Orissa Government to continue and expand the experiment in Santal Parganas and Orissa. An officer was employed in each of eight subdivisions, averaging nearly 1,000 square miles in extent, and in each some 15 centres were taken, so as to cover as widely as possible the subdivisional area. At each of these the officer was to spend four days, moving out on the first day two miles north, turning to the right and returning from the north-east, and on the succeeding days traversing the other three quadrants in the same manner. The aim was to distribute the sampling as regularly as possible both *in time* throughout the period of harvest and *in space* over the area sampled. The limitation in distance to be traversed secured that the number of samples obtained would be few in tracts where harvesting was sporadic and many in tracts where it was in full swing, thus correcting any tendency to over-emphasize the yields from relatively unimportant types. The instruction to take four samples from each field was withdrawn for reasons given below. For the supervision of this year's sampling I am indebted to Messrs. Mansfield and Houlton.

In all three years the actual process of taking each sample has been the same. The apparatus consists of a small solid block of wood, on which are fixed four pairs of guides defining two lines at an angle of 60° to one another. Along these lines through the guides are pushed two narrow wooden battens, which pass among the stalks of the paddy. Each bears a stop, so that, when this rests against the first pair of guides the length from the point where the battens cross to the slots near the ends of the battens is 67¼ inches. From one of these slots a third batten is pushed through the stalks to the slot on the other batten, and, when this third batten is fixed in the slots, an equilateral triangle is formed comprising an area of 1/3,200 of an acre. The paddy stalks just outside the battens are bent back, and the crop inside cut and threshed out there and then. The grain is placed in a numbered bag and dried for three or four days and then weighed. The size of the triangle has been chosen to give one maund an acre for every standard tola of the sample., while the triangular shape is intended to avoid the error caused by either including or excluding a whole line of the planted paddy. The weight can be accurately determined to half a tola in ordinary country scales by double weighing. The sampler is instructed to go to the centre of the *ail* (field boundary) roughly parallel to the line on which the harvest is proceeding and take three paces into the field before laying down the solid block of wood which guides the battens. He is warned against the mistake of putting the block down so that the triangle of battens encloses it instead of excluding it, a mistake to which beginners are liable.

The results obtained in the last three years are given in Tables I and II. Besides those to which reference has already been made, there are a set of samples taken in 1925 on the Kanke Agricultural Farm in Ranchi district and a set taken in the Atri thana of Gaya district in the same season The system has thus been tried in the Orissa deltaic tract, on the Chota Nagpur plateau of which Ranchi forms a part, in the broken country of Santal Parganas and Manbhum, and in the alluvial Gangetic plain of North Santal Parganas and Gaya. It has been put in practice in tracts, where the harvest has been plentiful and in others where it has been scanty. In 1925 the sampling extended in seven tracts over the whole period of the rice harvest, excluding only the very early *Bhadoi* and the summer rice, reaped in March, which are harvests of very little importance. Conclusions that can be drawn from the figures may, therefore,be applied with safety to the whole province except perhaps the North-Gangetic tract.

I have relegated to the Appendix the technical discussion of the validity of my conclusions, which will be appreciated only by those readers, who have made a somewhat extensive study of statistical theory. It is enough to say here that the mean yield on the harvested area can, I think, be determined for a tract of about 1,000 square miles with an accuracy such that it will be out by more than one maund per acre in only one case out of twenty. It will be very seldom out by more than a maund and a half. To secure such determination, it is necessary to fix 12 centres, spread as evenly as possible over the area, and to put down against each centre the day on which sampling is to be done. These days should be spread evenly over the period of harvest. The sampler should go out a fixed distance in one direction, and, circling round, return from another direction so that it is secured that he covers approximately the same area on each day. He should cut one sample from each field, where he finds harvesting in progress. He will be able ordinarily to get from 30 to 40 samples in a day when the harvest is in full swing and 10 to 20 when it is slack. It is certainly not impossible to supply samplers for this work, which is extremely simple once it is explained. The men employed hitherto have been young men fresh from college. Any gazetted officer and most other officers could do it quite well. In fact I have found that peons attached to the sampling officers fully understand what has to be done after a week or so. It is not in the least essential that the same officer should carry out all the sampling of a tract.

The method is applicable to small or large tracts. The same number of centres will probably give a slightly greater degree of accuracy for a smaller and a slightly less degree for a larger tract, because the standard deviation on the whole increases with the size of the tract, though not at all rapidly. Thus the standard deviation of the whole of the Santal Parganas is only 9.26 maunds, actually less than that of Deoghar, though greater than that of any other subdivision. To determine the mean yield of a village or an agricultural farm it would probably be advisable to make at least 10 cuts on each of 10 days. It would be better to make at least 20 cuts. (It may be noted that the standard deviation of Kanke in 1925 was nearly the highest figure found so far).

The method could also be applied to other crops, especially cereals. It would be more difficult to apply it to crops, for which the marketable produce is obtained by processes more elaborate than threshing, or is bulky and awkward to weigh, such as jute or sugarcane. It is also an advantage if the crop is common enough to enable a fair number of samples to be taken on one day.

It is no advantage to take a large number of samples from places very close together where the crops will naturally be very much the same on the same day. The degree of accuracy is not seriously improved by such practice. This explains why there is no need to take large samples instead of the handy samples obtained by my method. A sample of one-tenth of an acre is merely 320 of my samples taken in juxtaposition. It simply gives a determination of the mean yield of that particular field, which is not more effectively accurate than that given by say four small samples. Even four samples instead of one are not worth while, because in the great majority of cases they do not differ among themselves enough to affect the mean or the standard deviation of the whole set of samples. This may be illustrated from the columns for "all classes" and "1st cutting" for Santhal Parganas, Godda thana 1924. Technically speaking, there is very high correlation between the individuals of such groups of samples, which makes the ordinary rule, that the standard deviation of the mean is the standard deviation of the population divided by the square root of the number of samples, quite inapplicable.

A feature of all the frequency distributions obtained is their tendency to have a fairly level crest, the frequency not varying greatly for some four or five class intervals around the mean. This emphasizes the difficulty of "selecting an average field", the present method of experimenting. There will be nearly as many places where the crop is 4 maunds above or below the mean, as there will be where it is within one maund of it. Thus from the very nature of the rice crop, apart from the physical impossibility of surveying the whole of it simultaneously or nearly so, and the personal error of the experimenter, the present method of estimating is likely to be very far from accurate.

My method gives a reasonably close approximation for the mean yield of the *harvested area*. Consequently to get the gross yield it is necessary to determine what is that harvested area. It is evidently not the area planted or sown, since in practically all years some fields bear no crop worth cutting. It also excludes the field ridges (*ails*), which are usually measured in with the fields at a cadastral survey. Since the cadastral survey figure is the basis for the cropped area of the year, this second point is of importance. No very satisfactory determination of the area occupied by ridges has yet been made, but it would appear to vary between nearly 10 per cent in hilly country to scarcely more than 2 per cent in level country. It is usually taken as 5 per cent.

It is much more difficult to say what is the range of variation between the planted (or sown) area and the harvested area. In a bad year, when the rains stop early in September, it may well be 50 per cent all over a subdivision. The planted area will also vary from year to year. At present we depend on estimates, which are originally those of the village chaukidar though they are much more capable of intelligent check by higher authorities than estimates of yields are. It is possible, for instance, by driving through a tract in a motor car to get some idea of the extent to which the rice crop is going to fail altogethr. It is usually believed that the estimate of cropped area is sufficiently accurate. It may be, but the belief is intuitive rather than rational.

It would certainly be a great advantage if the harvested area could be obtained by some form of sampling. A method was tried in 1925 in Santal Parganas. The results, if not very encouraging, are interesting. The plan was to make the sampler march from centre to centre across country as near as he conveniently could in a straight line. After certain intervals of time he had to count 100 paces and note how many of these ended on harvested rice land. (There is no serious difficulty in doing this in January and early February, since little rice land is reploughed by then). Each count of 100 would give a definite percentage, in many caes 0, in many nearly 100, the difference from the full 100 being due to the presence of field ridges. The mean of these percentages would give the percentage of harvested rice land in the total area of the tract sampled. Six men were employed for nearly a month each and between them made 1971 counts. From these a mean of 47.22 per cent. was obtained. The standard deviation was 37.10. This means that it is about 21 to 1 that the true percentage is between 49.4 and 45. The accuracy obtained is hardly sufficient to justify the employment of six men for a month, in each district. Further the work is strenuous and tedious, and it is probable that it would be in practice shirked, ard results fudged. The difficulty of excluding large patches of thick jungle or other inaccessible country both from the sampling and the sampled area has yet to be overcome. But it is still possible that some method on similar lines may prove practicable.

There is just one other caution which I would address to persons who have to estimate gross yields. That is that yields of rice are ultimately based, however imperfectly at present, on yields of paddy. A standard figure for the loss of weight by husking is taken, usually one-third. It is probable that this varies a good deal with the kind of paddy, and there is good reason to believe that in Orissa the loss is considerably more than one-third. It would be well if the Agricultural Department investigated this point on their farms. If any kind of paddy husks out much below the usual, it is clearly not such a good economic proposition, as an excess, of unhusked yield might suggest. Possibly the plant breeders may be able to produce a paddy, which loses less than one-third in husking, and, if so, the benefit might well be as great as that obtained from increased yield.

My general conclusions are that sampling by my method on 12 days throughout the harvest from different centres would secure a figure for the mean yield of a harvested area of 1,000 square miles correct within one maund in about 95 per cent. of cases, and that with the existing revenue and agricultural staff there would be no serious difficulty in substituting this method for the unscientific method of sampling now prescribed.

To the officers I have mentioned, under whose directions the samples were taken, and to the staff employed by them on the work I offer my cordial thanks as I do to Mr. Dobbs for his advice and encouragement.

APPENDIX

It is often assumed that if a number (n) of samples are taken and the standard deviation of the whole population (σ_0) is known, then the standard deviation (σ_m) of the mean of n samples is given by $\sigma_m{}^2=\sigma_0{}^2/n$. This however, presupposes (1) that the samples are taken by the process of "simple" or "random" sampling, i.e., that every portion of the population is equally likely to be sampled every time, and (2) that the distribution of the population is such that the distribution of the means of n samples closely approximates to the "normal" curve of error.

I have attempted to investigate both these points before recommending the method already tried for general application. The application of "simple" sampling to the rice crop would involve picking out the places for sampling by some random method such as putting numbered cards representing all the rice plots of the area in a box and drawing out the required number of cards, and then arranging to visit the plots at the time when the crop is ripe. Such a method is, I consider, impracticable. My method is practicable but is not "simple" sampling. There is some degree of selection. It has to be seen whether the adoption of it in place of "simple" sampling increases or decreases the rate at which the standard deviation of the mean diminishes as the number of cuts rises.

If an equal number of cuts were taken on each day, say 20, the process would be equivalent to (1) taking one cut from each separate tract to make up a sample, and then (2) taking the mean of 20 such samples ; the "tract" being the country which the sampling officer can cover from a centre. Process (1) theoretically gives a set of samples with means having a standard deviation given by $\sigma_{m_1}^2=\left(\sigma_0^2-S_m^2\right)/n_1$ where σ_0 is again the standard deviation of the population, S_m that of the mean values of the tracts and n_1 the number of the tracts. Process (2) is simple sampling of the means of the samples obtained by process (1), and will give a mean with a standard deviation given by $\sigma_{mk}^2=\sigma_{m_1}^2/k$ where k is the number of samples grouped together (in this case 20).

To test the theory I have taken from the original record of Rajmahal 1925 the first eight cuts of each day, writing the 32 cuts of the first four days in one row, the 32 cuts of the second four days in the next row, and so on. (In a very few cases, where less than eight cuts were taken in a day, I have borrowed from the last cuts of the previous day or the following day). I have thus obtained 32 columns with 11 figures in each corresponding to the 11 centres, each surrounded by four quadrants. I have totalled the columns and taken the means. I have then taken the means of successive pairs, fours, and eights of the means so derived. I have also calculated S_m^2 by taking the means of the cuts made from each centre. I have done the same thing for Bhadrak 1925, which had 12 centres instead of 11. The results are as follows (standard deviation in maunds).

		σ_0^2	S_m^2	$\sigma_{m_1}^2$	$\sigma_{m_2}^2$	$\sigma_{m_4}^2$	$\sigma_{m_8}^2$
Rajmahal	theoretical	69·06	34·21	3·17	1·58	·79	·40
,,	observed	—	—	1·62	1·08	·40	·31
Bhadrak	theoretical	33·18	9·46	1·98	·99	·50	·25
,,	observed	—	—	1·10	·74	·54	·11

(The theoretical value of $\sigma_{m_1}^2$ is $\sigma_{m_0}^2 - S_m^2$ divided by 11 for Rajmahal and by 12 for Bhadrak. The theoretical values of $\sigma_{m_2}^2$, $\sigma_{m_4}^2$ and $\sigma_{m_8}^2$ are one half, one quarter and one-eighth of $\sigma_{m_1}^2$)

The observed results agree sufficiently well with the theory, and it seems quite clear that the standard deviation is diminished and not increased by taking samples on this plan instead of by simple sampling.

But this plan would require an equal number of cuts on each day, and this would certainly overweight the slack periods of harvest. My method avoids this, but by so doing it almost certainly increases the standard deviation of the mean as compared with that obtained by the plan just described. I cannot find any way of determining mathematically to what extent it does so, but the following test shows reasonably well what happens.

I examine below the means obtained from the cuts made to the north-east of each centre against those obtained from the cuts made to the south-east, south-west and north-west. These are given below. (Some cuts used in the general table were neglected for this purpose, because only one quadrant was sampled).

——	average no. of cuts in a quadrant	1st. quadrant	2nd quadrant	3rd. quadrant	4th. quadrant	all	standard deviation	
							actual	theoretical
All S. P.	1432¼	16·47	16·85	17·32	17·42	17·02	·25	·25
Deoghar	280	20·01	19·12	21·31	20·23	20·05	·72	·69
Dumka	122	19·57	20·52	19·46	20·63	20·06	·44	·71
Godda	342¼	12·64	14·05	15·05	15·13	14·25	·65	·35
Jamtara	216	13·60	14·39	12·09	13·43	13·41	·78	·52
Rajmahal	312¼	21·08	21·38	22·06	22·30	21·67	·37	·47
Pakaur	159¾	10·95	10·80	11·69	11·11	11·07	·19	·52
Bhadrak	288¾	13·13	14·60	13·52	13·29	13·65	·49	·34

The "failures", i.e., results falling more than one maund away from the mean of the four quadrants are underlined. The actual standard deviation is that derived from the means of the four quadrants, Sheppard's correction being applied. The theoretical standard deviation is the result of dividing the standard deviation of the whole series of cuts in the area by the square root of the average number of cuts in the quadrants. The "failures" look high at first sight, but that is due to the fact that the number of cuts in a quadrant was often distinctly lower than the standard of 20 cuts at 12 centres, which would seem to be what is required to ensure a mean correct within one maund in 95 per cent of cases. A comparison of the actual with the theoretical standard deviation shows that on the whole my method of sampling gives about as stable a mean as simple sampling. The variation in the number of cuts at the different centres about wipes out the greater stability gained by distributing the cuts to the different centres.

It is impossible to demonstrate by mathematical processes precisely what the odds are, and, without a much wider series of tests than have yet been made empirical deductions must be somewhat uncertain. But I think it is quite safe to say that if at least 10 and not more than 40 cuts were made from each of the 12 centres at even intervals over the harvest season, the mean yield for an area of about 1000 square miles could be determined within one maund per acre nineteen times out of twenty. The "failures" would very seldom be more than 1½ maunds out. This depends on the assumption that the frequency distributions actually found are such that the means of a large number of cuts will form a frequency distribution closely approximating to the "normal curve of error". Where the sampling is "simple", this is certainly true of populations following the frequency distributions obtained. If, for example, 400 samples were taken by "simple" sampling from population like the Deoghar 1925 and the Godda 1925 harvests, the spread of the mean of the 400 samples would compare with the spread of the corresponding "normal" curve (i.e., that having the same standard deviation) as follows.

limits			Deoghar	Normal
± ·425	mds	..	1125 inside to 1000 outside	1138 inside to 1000 outside
± ·850	,,	..	585 ,, to 100 ,,	583 ,, to 100 ,,
±1·275	,,	..	364 ,, to 10 ,,	330 ,, to 10 ,,
±1·700	,,	..	342 ,, to 1 ,,	272 ,, to 1 ,,
			Godda	Normal
± ·3428	,,	..	2517 inside to 1000 outside	2531 indide to 1000 outside
± ·5142	,,	..	822 ,, to 100 ,,	837 ,, to 100 ,,
± ·6856	,,	..	3011 ,, to. 100 ,,	3050 ,, to 100 ,,
± ·8570	,,	..	1357 ,, to 10 ,,	1375 ,, to 10 ,,

It cannot then be said with certainty, what effect on the rapidity of the approach to normality would arise from the substitution of my method of sampling for "simple" sampling. But as it has little or no effect of diminishing or increasing the standard deviation of the mean, it probably has the effect of bringing the higher constants about as quickly towards their values for the normal curve. The nearest approach to a discussion of this problem is to be found in a paper by "Student" in Vol. VII pages 210 to 214 Biometrika, but the problem before him was not precisely mine, and he distinctly states that his problem is essentially a "small sample" problem. It deals with means of 25, not as mine with means of 250 or so. The higher constants for some of the distributions are given below (Professor Pearson's notation).

				β_1	β_2
Godda	1925	..	..	·527	4·116
Deoghar	1925	..	..	1·227	4·066
Pakaur	1925	..	..	1·348	6·349
Godda	1924 1st cutting			·030	2·729
Jaipur	1924 1st cutting			·694	3·388

Professor Bowley's tables on pages 271 and 303 of his "Elements of Statistics" enable one to show how the observed mean is likely to be about the true mean. Assuming 256 samples for each tract, 10,000 trials would be distributed thus :

(κ is Professor Bowley's measure of skewness and is μ_3/σ^3 in Professor Pearson's notation)

——	σ	(κ)	not more than 1 md. above mean	not more than 1 md. below mean	more than 1 md. from mean
Godda 1925	6·63	·64	4,888	4,956	156
Jamatra 1925	7·70	·78	4,752	4,844	404
Pakaur 1925	6·54	·86	4,900	4,982	118
Rajmahal 1925	8·31	·26	4,712	4,752	536
Deoghar 1925	11·59	1·11	3,999	4,225	1,776
Dumka 1925	8·15	·59	4,706	4,782	512
Jaipur 1925	9·09	1·51	4,536	4,714	750
Bhadrak 1925	5·76	·69	4,943	5,003	54

It may, of course, be argued that my method, even if the same or nearly the same mean were obtained in a large number of repetitions, may be defective owing to bias. That must remain a matter of opinion. The method does, however, secure that the cutting is well distributed in space over the area to be sampled and in time over the period of the harvest. By laying down that approximately the same area shall be under sample on each of the days of sampling, it weights the importance of the contribution made to the mean by the harvest of the day. A day in the height of the harvest will contribute 30 to 40 samples; in the slacker time only 10 to 20. There is no other obvious form of bias, unless of course, centres are selected with grave disregard to their suitability.

If it be admitted that the distribution of the means of a fairly large number of cuttings will closely approximate to a "normal curve of error" and that the standard deviation of the means is approximately that of the population divided by the square root of the number of cuttings, then the degree of accuracy can be calculated. Leaving aside the standard deviations of rice grown on particular classes of land, which are sometimes very low, it may be taken that the range of the standard deviation of rice in Bihar and Orissa is from 4 to 12 maunds. At the lower limit, 240 cuttings would give means of which only 15 out of 100,000 would be one maund or more away from the true mean. At the upper limit about 19 out of 100 would be so divergent, and about 5 out of 100 would be more than 1½ maunds away. With a standard deviation of 8 maunds, 240 cuttings would give a mean correct within one maund about 95 times out of 100. This last is the basis of my scheme of 12 centres, from which may be expected at least 240 cuttings, but probably as many as 300. Where experience shows that the standard deviation of a tract is much higher than 8 maunds, the number of centres should be increased, if it is desired to obtain the same degree of accuracy.

It is of interest to consider what degree of accuracy would attach to a calculation of the gross yield of the whole province of Bihar and Orissa, based on the system I recommend, the harvested area being estimated as it is present. If it may be assumed that the standard deviation of the error in estimating the harvested area is about 1/20 of that estimate, it is practically certain (1,000 to 3) that the calculation of the gross yield is not more than 2 per cent. out.

TABLE 1. FREQUENCY DISTRIBUTION OF RICE YIELDS 1923-24

		Santal Parganas, Godda Thana, 1924					Manbhum, Purulia Thana, 1924					Cuttack, Jaipur Thana 1924	
class interval in maunds	Santal Parganas Godda Thana 1923 all classes,	1st class Lands	2nd class Lands	3rd class Lands	all classes	1st cuttings	1st class Lands	2nd class Lands	3rd class Lands	all classes	1st cuttings	all	1st cuttings
1	2	3	4	5	6	7	8	9	10	11	12	13	14
0— 2⅛	55	12	1	26	39	8	—	1	2	3	1	28	16
2⅛— 4¼	24	28	24	48	100	28	1	2	5	8	2	68	37
4¼— 6⅜	44	69	49	82	200	43	6	8	40	54	18	102	45
6⅜— 8½	45	128	53	90	271	71	17	11	74	102	33	124	43
8½—10⅝	71	126	86	47	259	82	18	17	84	119	37	115	42
10⅝—12¾	27	175	78	30	283	62	18	18	100	136	42	124	52
12¾—14⅞	18	223	59	60	342	79	17	31	78	126	39	83	34
14⅞—17	21	192	47	23	262	67	22	28	77	127	50	87	37
17 —19⅛	19	173	27	4	204	57	31	19	70	120	40	52	14
19⅛—21¼	22	118	12	4	134	49	13	27	33	73	27	39	21
21¼—23⅜	15	51	9	3	63	17	18	16	22	56	23	24	5
23⅜—25½	14	16	6	3	25	4	16	11	13	40	20	28	11
25½—27⅝	12	4	3	—	7	1	10	8	9	27	15	21	11
27⅝—29¾	4	6	6	—	12	3	10	5	4	19	10	17	4
29¾—31⅞	1	2	2	—	4	1	7	8	2	17	10	8	5
31⅞—34	4	6	—	—	6	3	4	3	—	7	3	4	1
34—36⅛	2	—	—	—	—	—	—	—	—	—	—	2	1
36⅛—38¼	0	—	—	—	—	—	1	4	—	5	2	2	0
38¼—40⅜	1	—	—	—	—	—	—	—	—	—	—	1	1
40⅜—42½	0	—	—	—	—	—	—	—	—	—	—	1	—
42½—44⅝	1	—	—	—	—	—	—	—	—	—	—	—	—
44⅝—46¾	—	—	—	—	—	—	—	1	—	1	—	—	—
46¾—48⅞	—	—	—	—	—	—	1	—	—	1	—	—	—
total no. of cuttings	400	1,329	462	420	2,211	575	210	218	613	1,041	372	930	380
mean in mds	10·67	13·83	11·74	8·56	12·70	12·63	17·76	16·37	13·40	15·59	16·03	12·25	11·81
standard deviation in mds.	8·06	5·30	5·38	4·71	5·60	5·68	7·48	7·50	5·36	6·59	6·91	7·13	7·20
* c. v.	75·5	38·3	45·8	55·0	44·1	45·0	42·1	45·8	40·0	42·3	43·1	58·2	61·0
period of cuttings	4th Dec.—22nd Dec.	14th Nov. 13th Jan.	10th Nov.—27th Dec.	10th Nov.—19th Dec.	10th Nov.—13th Jan.	10th Nov.—13th Jan.	29th Nov.—22nd Dec.	17th Nov.—19th Dec.	16th Nov.—13th Dec.	16th Nov.—22nd Dec.	16th Nov.—22nd Dec.	17th Dec.—2nd Jan.	17th Dec.—2nd Jan.

---

* Coefficients of variation have been inserted by us—*Editor.*

Table 2. *Frequency ditribution of rice yields*, 1925.

class interval in mds.	Godda Thana (Santal Parganas)				Jamtara Thana				Pakaur Thana				Rajmahal Thana			
	1st.	2nd.	3rd.	all	1st.	2nd.	3rd.	all	1st.	2nd.	3rd.	all	1st.	2nd.	3rd.	all
0— 2⅛	0	0	8	8	0	1	42	43	0	2	18	20	0	0	0	0
2⅛— 4¼	1	5	46	52	1	9	44	54	1	13	36	50	0	0	7	7
4¼— 6⅜	2	17	73	92	2	18	36	56	3	36	64	103	0	0	24	24
6⅜— 8½	0	42	72	114	14	41	31	86	4	42	44	90	0	1	26	27
8½—10⅝	12	90	54	156	20	61	26	107	15	40	39	94	0	19	42	61
10⅝—12¾	21	115	34	170	31	52	20	103	12	26	28	66	0	84	12	96
12¾—14⅞	46	97	35	178	29	54	6	89	11	27	13	51	0	96	—	96
14⅞—17	64	91	17	172	24	47	2	73	12	20	4	36	0	99	—	99
17—19⅛	72	66	9	147	42	34	1	77	12	24	6	42	37	73	—	110
19⅛—21¼	58	35	3	96	25	21	—	46	12	24	2	38	120	18	—	138
21¼—23⅜	37	17	1	55	15	13	—	28	13	8	1	22	141	10	—	151
23⅜—25½	23	15	—	38	18	13	—	31	5	5	0	10	133	—	—	133
25½—27⅝	25	3	—	28	16	9	—	25	1	2	1	4	74	—	—	74
27⅝—29¾	15	4	—	19	9	4	—	13	3	3	—	6	54	—	—	54
29¾—31⅞	15	5	—	20	11	2	—	13	2	1	—	3	49	—	—	49
31⅞—34	4	3	—	7	3	1	—	4	1	—	—	1	82	—	—	82
34 —36⅛	5	2	—	7	3	2	—	5	1	—	—	1	45	—	—	45
36⅛—38¼	3	—	—	3	1	2	—	3	0	—	—	0	19	—	—	19
38¼—40⅜	2	—	—	2	0	—	—	0	0	—	—	0	23	—	—	23
40⅜—42½	—	—	—	—	2	—	—	2	0	—	—	0	10	—	—	10
42½—44⅝	—	—	—	—	1	—	—	1	0	—	—	0	3	—	—	3
44⅝—46¾	—	—	—	—	0	—	—	0	0	—	—	0	2	—	—	2
46¾—48⅞	—	—	—	—	1	—	—	1	0	—	—	0	—	—	—	—
48⅞—51	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—
51 —53⅛	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
53⅛—55¼	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
55¼—57⅜	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
57⅜—59½	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
59½—61⅝	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
61⅝—63¾	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
total no. of cuttings	405	607	352	1,364	268	384	208	860	109	273	256	638	792	400	111	1,303
mean in mds.	19·61	14·13	8·50	14·30	18·40	13·95	5·87	13·38	16·87	12·14	7·59	11·12	26·56	14·95	7·97	21·41
standard deviation in mds.	5·96	5·15	4·20	6·63	7·47	6·20	3·84	7·70	7·42	6·31	4·26	6·54	4·97	2·94	2·34	8·31
c. v.	30·4	36·4	49·4	46·4	40·6	44·5	65·4	57·5	44·0	52·0	56·1	58·8	18·7	19·7	29·4	38·8
period of cuttings	Oct. 20th-Jan. 1st				Oct. 20th-Dec. 17th				Oct. 21st-Dec. 20th				Nov. 5th-Jan. 2nd.			

Table 2 (contd.) Frequency distribution of rice yields 1925

class intervals in mds	Deoghar Thana				Dumka Thana				Jaipur Thana (Cuttack)	Kanke Farm (Ranchi)	Atri Thana (Gaya)	Bhadrak Thana
	1st	2nd.	3rd	all	1st	2nd.	3rd.	all	all			all
$0—2\frac{1}{8}$	0	0	2	2	0	0	0	0	31	0	6	2
$2\frac{1}{8}—4\frac{1}{4}$	0	1	13	14	0	0	3	3	64	1	8	36
$4\frac{1}{4}—6\frac{3}{8}$	1	4	47	52	0	1	13	14	79	1	15	68
$6\frac{3}{8}—8\frac{1}{2}$	10	10	60	80	0	3	16	19	79	1	32	131
$8\frac{1}{2}—10\frac{5}{8}$	16	28	58	102	0	9	16	25	79	2	21	142
$10\frac{5}{8}—12\frac{3}{4}$	25	33	42	100	5	10	13	28	68	2	26	138
$12\frac{3}{4}—14\frac{7}{8}$	17	37	19	73	8	18	18	44	52	4	14	186
$14\frac{7}{8}—17$	39	40	19	98	19	22	8	49	60	5	49	171
$17—19\frac{1}{8}$	32	39	21	92	30	14	20	64	41	6	21	99
$19\frac{1}{8}—21\frac{1}{4}$	48	35	14	97	20	22	7	49	39	3	25	75
$21\frac{1}{4}—23\frac{3}{8}$	31	22	10	63	25	15	5	45	29	3	23	35
$23\frac{3}{8}—25\frac{1}{2}$	33	11	5	49	26	15	5	46	17	2	30	27
$25\frac{1}{2}—27\frac{5}{8}$	34	10	7	51	18	6	6	30	7	6	23	18
$27\frac{5}{8}—29\frac{3}{4}$	39	14	2	55	12	0	1	13	8	6	2	12
$29\frac{3}{4}—31\frac{7}{8}$	19	5	—	24	13	0	1	14	2	7	14	5
$31\frac{7}{8}—34$	26	11	—	37	16	0	0	16	6	4	5	4
$34—36\frac{1}{8}$	27	4	—	31	10	2	2	14	4	6	3	2
$36\frac{1}{8}—38\frac{1}{4}$	8	5	—	13	2	1	—	3	5	3	2	1
$38\frac{1}{4}—40\frac{3}{8}$	16	0	—	16	2	—	—	2	5	1	2	1
$40\frac{3}{8}—42\frac{1}{2}$	15	1	—	16	3	—	—	3	3	2	4	—
$42\frac{1}{2}—44\frac{5}{8}$	6	—	—	6	5	—	—	5	0	1	1	—
$44\frac{5}{8}—46\frac{3}{4}$	9	—	—	9	2	—	—	2	2	2	—	—
$46\frac{3}{4}—48\frac{7}{8}$	9	—	—	9	0	—	—	0	3	1	—	—
$48\frac{7}{8}—51$	6	—	—	6	1	—	—	1	0	2	—	—
$51—53\frac{1}{8}$	5	—	—	5	—	—	—	—	2	0	—	—
$53\frac{1}{8}—55\frac{1}{4}$	6	—	—	6	—	—	—	—	2	1	—	—
$55\frac{1}{4}—57\frac{3}{8}$	5	—	—	5	—	—	—	—	—	—	—	—
$57\frac{3}{8}—59\frac{1}{2}$	6	—	—	6	—	—	—	—	—	—	—	—
$59\frac{1}{2}—61\frac{5}{8}$	5	—	—	5	—	—	—	—	—	—	—	—
$61\frac{5}{8}—63\frac{3}{4}$	1	—	—	1	—	—	—	—	—	—	—	—
total no of cuttings	494	310	319	1,123	217	138	134	489	687	72	326	1,153
mean in mds	27·07	18·13	11·29	20·12	24·64	18·08	14·21	19·93	12·88	26·63	17·46	13·63
standard deviation in mds.	11·71	7·25	5·76	11·59	7·69	5·57	6·80	8·15	9·09	11·30	8·73	5·76
c. v.	43·3	40·0	51·0	57·6	31·2	30·8	47·9	40·9	70·6	42·4	50·0	42·3
period of cuttings	Oct. 20th—Dec. 26th.				Oct. 25th—Dec. 21st.				Nov. 23rd Dec. 31st	Dec. 4th Jan. 1st		Oct. 23rd Jan. 11th

Paper received 10 November 1945.

CONFOUNDED FACTORIAL DESIGNS IN QUASI-LATIN SQUARES

By C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta.

INTRODUCTION

In factorial experiments the number of treatments become large with increase in the number of factors and also in the number of levels of the factors, so that randomized block designs cannot be used with much efficiency. To combat this all the treatment comparisons are assigned to two or more blocks. By so doing the comparisons arising from the groups of treatments assigned to various blocks cannot be isolated from the fertility differences arising from the blocks. In such a case, the above comparisons are said to be confounded with the blocks. In two or more replications, different sets of comparisons may be confounded in different sets of blocks. But if we choose a Latin square arrangement and arrange the several treatments such that neither should a treatment occur more than once in a row or a column nor all treatments occur in a row or a column, some comparisons will be confounded in the rows and some in the columns. Such arrangements are called Quasi-Latin squares (Yates, 1937).

2. THE GENERAL PROBLEM OF DOUBLE CONFOUNDING.

Let the number of plots in a row or a column be k and kn the total number of treatment combinations. If $k=nr$, then with r replications a latin square arrangement can be found in which some comparisons are confounded with the rows and some with the columns. In order that our experiment may not suffer from the defect of causing apparent effects due to the interaction of fertility and some of the factors it is always better in factorial experiments to group together all the blocks that form a complete replication. This imposes a further restriction on the design that the rows and columns can be broken into r sets each forming a replication with n blocks. For the actual layout we have to randomise the replications, both column and row wise, and then randomise the columns and rows within the replications.

The analysis of such designs does not present any fresh difficulty. We give here the general scheme of analysis in case in which there are m latin square arrangements of size $k \times k$ with r replications in each, confounding some effects with the rows and some with the columns. The rows and columns each have $m(k-1)$ degrees of freedom and variances due to them can be obtained from the row and column totals. The unconfounded interactions and main effects can be calculated from the treatment totals obtained from all the squares. Due to the property of orthogonality of the Latin square it is clear that the effects confounded with the rows (or columns) cannot be estimated from comparisons within the columns (or rows). Another property which gives the clue to the analysis is that any confounded interaction or main effect with the rows (or columns) in a replication can be estimated from other replications constituted by the rows (or columns) in which it is not confounded. This gives a rule which is similar to the rule of finding the sum of squares due to the partially confounded effects ; we need choose the treatment totals only from those replications in which the particular effect is not confounded either in the rows or columns . This is at once detected by the two rules, stated above, regarding the estimability of an effect. The loss of information on the partially confounded effects is calculated in the same way as in the randomized incomplete block designs. Loss of information on any effect is one minus the ratio of the number of replications from which the effect is estimable to the total number of replications.

3. CONSTRUCTION OF DESIGNS WITH DOUBLE CONFOUNDING

The next problem is to enunciate the combinational problem and discuss the possibilities of the actual construction of designs. The case of symmetrical factorial confounded designs will be discussed first and some of the results are extended to the asymmetrical case.

Let there be 'a' factors each at s levels. If $k = s^b$ then the relationships $n\,k = s^a$, $nr = s^b$ and $1/r = s^{a-2b}$ hold. Evidently $a \leqslant 2b$. The values of $r = 1, 2, 3$ and 4 only give rise to designs which are of practical value. The useful values of r, s and b are listed below.

Table 1

r	s	b
1	prime or prime power	$a/2$
2	2	$(a+1)/2$
3	3	$(a+1)/2$
4	2	$(a+2)/2$
4	4	$(a+1)/2$

In the case of $r = 1$, the construction of designs can be very simply effected by first forming the design confounding certain effects in the rows and adjusting the columns such that an orthogonal set is confounded in them. The adjustment is facilitated by the property that each of the columns confounding an orthogonal set can be built up by chosing one element from each row.

The introduction of the geometrical method of getting at confounded designs, developed by Bose and Kishen (1941) when s is a prime or a prime power may be used at this stage. They have shown that s^{a-b} blocks confounding $(s^{a-b}-1)$ degrees of freedom can be identified with the bundle of flats each of b dimensions passing through a common $(b-1)$ flat at infinity. The nature of confounding depends on the way in which all the $(a-2)$ flats passing through the above $(b-1)$ flat at infinity stand in relation to the 'fundamental simplex' defined by them. From this we have the important deduction that two confounded systems in s^b plots blocks have their effects orthogonal if and only if their corresponding flats at infinity do not have a common $(a-2)$ flat at infinity. In the case $b = a/2$ (a is necessarily even), it is always possible to find $(a/2-1)$ dimensional flats such that they don't have any $(a-2)$ flat in common, but the best designs are only those in which the highest order interactions are confounded. After choosing a suitable pair the arrangement is supplied by the points of interesection of the bundle of flats emanating from the two $(b-1)$ flats at infinity, one set of flats corresponding to the rows and another to the columns.

With more than one latin square, different sets of effects can be confounded in the rows and columns. Sometimes it is possible to get balanced designs with half the number of replications necessary in the case of randomised block designs. We shall discuss only those designs in which $k \leqslant 10$ for only these will be of practical value.

If $s = 2$, $a = 4$, and $b = 2$, there are four factors A, B, C, and D and 3 degrees of freedom are confounded with the rows and 3 with the columns.

(a) Using two latin squares of size 4 × 4 we can confound the sets

AB ; CD ; ABCD.
AD ; ABC ; BCD.

in the columns and rows of one latin square and the sets

AC ; BD ; ABCD.
BC ; ABD ; ACD.

with those of the other. The relative information on the first and second order interactions is $\frac{1}{2}$ while the third order interaction is completely confounded.

(b) We may use the following sets

AB ; ACD ; BCD.
CD ; ABC ; ABD.

BC ; ABD ; ACD.
AD ; ABC ; BCD.

AC ; BCD ; ABD.
BD ; ACD ; ABC.

in three latin squares and get a balanced arrangement giving $\frac{2}{3}$ of the relative information on the first order interactions and completely confounding the second order interactions.

(c) With five latin squares we can confound the sets

AC ; BD ; ABCD.
AD ; ABC ; BCD.
BC ; ABD ; ACD.

two by two in three latin squares and the sets

A ; B ; AB.
C ; D ; CD.

repeatedly in two latin squares. The relative information on all the effects is $\frac{3}{5}$. If we ignore the gain in efficiency due to arrangement in a latin square, the randomised block design with five replications confounding the above sets is better than the Latin square arrangement in that the relative information on each effect is $\frac{4}{5}$.

In the case $s = 2$, $a = 6$, $b = 3$, there are six factors A, B, C, D, E and F each at two levels. There is no simple balanced design in less than five replications. Unbalanced designs can always be obtained in one or more latin squares by confounding orthogonal sets of effects in the rows and columns.

In the case, $s = 3$, $a = 4$, $b=2$, it has been shown by Yates (1937) and Nair (1938) that there exists a design in four replications with randomised blocks which gives balance to the partially confounded second order interactions. The fact that the confounded effects in several replications are orthogonal facilitates the balanced arrangement in two Latin squares. The layout is obtained by merging the replications two by two as indicated in para 4. of § 3. The relative information on the partially confounded interactions in $\frac{1}{2}$, where as in the balanced randomised block design is $\frac{3}{4}$. But the Latin square arrangement saves the cost of two more replications. By replicating the two latin squares the design becomes as efficient as the corresponding randomised block design with equal number of replications, even if we ignore the fact that the variance per plot is reduced in the latin square arrangement.

It is not worthwhile discussing the higher types of designs from the practical point of view. But we may give the general result that if in s^a design with s^b plots blocks $(b = a/2)$ balance can be obtained in randomised blocks with $2r$ replications and the effects that are confounded in them can be split into r pairs of orthogonal sets, then balance can be achieved in r latin squares. If $2r+1$ replications are necessary in randomised blocks then utmost $2r+1$ latin squares are required. But these will not be as efficient as the randomised block designs.

We shall investigate the possibilities of constructing designs in the case $r>1$ and derive the necessary and sufficient conditions in particular cases. Whenever a design with conditions stated in para. 1 of Section 2 exists then it is always possible by rearrangement to adjust such that the first, second etc., sets of rows and columns form complete replications. By adjustment the treatments in the quasi-latin square form into r^2 groups of n^2 treatments each, the groups coming in the same row or column not having any element in common. These groups need not necessarily be identical in sets of r, but this holds automatically in the case $r=2$. If we take any n columns forming a complete replication together with the division into r groups as formed above we get an arrangement in blocks of size k/r confounding $(nr-1)$ degrees of freedom. In these the $(n-1)$ degrees of freedom confounded in the n columns are necessarily involved. Dealing with columns only we find r such replications in blocks of size k/r confounding $(nr-1)$ degrees of freedom each. In order that the effects confounded in any replication of the rows may be estimated completely from the replications in which it is not confounded, it is necessary that any row of the latin square should be capable of being built up by choosing one element from each of nr blocks of size k/r of any replication of the columns. This shows that the effects to be confounded in the rows should be orthogonal to the totality of the confounded effects in the various replications obtained from the columns in blocks of size k/r. A similar result holds for the effects confounded in the columns and the two results are complementary.

In the case $r=2$, there are four groups and the treatment combinations in the diagonal groups are the same. If we consider any replication of the columns, confounding $(n-1)$ degrees of freedom, then the effect with one degree of freedom, confounded in the two groups into which it is split in the manner indicated above, together with $(n-1)$ degrees of freedom determines the totality of the $(2n-1)$ d.f. confounded in a replication with blocks of size $k/2$. Similarly the same effect confounded with the groups in the other replication of the columns determines the totality of the $(2n-1)$ degrees of freedom confounded in blocks of size $k/2$. The effects that are confounded with the rows must be orthogonal to these different effects confounded above. So the necessary and sufficient condition for the existence of a quasi-latin square design with two replications is that there exists a common effect which together with effects confounded in the columns (or rows) and their interactions with it are orthogonal to the effects to be confounded in the rows (or columns). The number of different solutions will depend on the number of such available effects. The actual construction of some designs are given below.

If $s=2$, $a=5$, $b=3$, there are 5 factors A, B, C, D and E each at 2 levels. The higher level may be indicated by the small letter and lower level by the absence of a letter. It is desired to confound the effects.

$$\begin{array}{lll} ABC; & ADE; & BCDE. \\ ABD; & BCE; & ACDE. \end{array} \qquad \ldots \quad (3.3)$$

with the rows and the effects

$$\begin{array}{lll} ACE; & BCD; & ABDE. \\ ACD; & BDE; & ABCE. \end{array} \qquad \ldots \quad (3.31)$$

with the columns. It is found that there are two effects ABCD and ABCDE which satisfy the conditions given above. The solution of the above problem given by Yates (1937) is that corresponding to ABCD. The alternative solution is given below. ABCDE in conjunction with (3.3) gives

ABC ; ADE ; BCDE ; ABCDE ; A ; BC ; DE.
ABD ; BCE ; ACDE ; ABCDE ; B ; AD ; CE.

Two replications are formed in blocks of 4 plots and arranged as indicated below.

TABLE 2. QUASI-LTAIN SQUARE FOR 2^5 DESIGN.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
R_1	abcde	ade	cd	bd	ce	a	be	abc
R_2	ad	abcd	bde	cde	ae	c	abce	b
R_3	abde	ac	e	bce	d	acde	bcd	abde
R_4	ace	abe	bc	1	acd	de	abd	bcde
R_5	de	b	abd	ae	abcde	bce	ac	cd
R_6	be	d	a	abde	bc	abcd	cde	ace
R_7	c	bcde	abce	acd	abde	bd	ade	e
R_8	bcd	ce	acde	abc	bde	abe	1	ad

The comparisons (3.3) are confounded in the rows R_1 to R_4 and R_5 to R_8 and ABCDE is confounded in the groups differentiated above. Now there remains only shifting of the treatments in each group without disturbing the rows, so that the comparisons (3.31) are confounded in the columns C_1 to C_4 and C_5 to C_8. The design given above is arranged in this manner and gives a suitable layout only after proper randomisation discussed in introduction. The relative information on the partially confounded effects is $\frac{1}{2}$. We may arrive at a balanced arrangement in 5 latin squares by confounding in each four sets of the five sets of comparisons obtained by adding : .. (3.4)

ABE ; CDE ; ABCD.

to (3.3) and (3.31). The relative information on each of the confounded interactions is only $\frac{3}{5}$. Designs confounding any four of the sets.

AB ; CD ; ABCD.
AC ; DE ; ACDE.
AD ; BE ; ABDE.
AE ; BC ; ABCE.
BD ; CE ; BCDE.

are also possible but may not be of use for the first order interactions are partially affected.

If $s = 2, a = 3, b = 2$, there are three factors A, B and C each at two levels. Various types of confounding and the common effects with whose help the designs can be formed are given in Table 3.

In the general case $r > 2$ we may state the necessray and sufficient condition as follows :

(a) If we denote the effects to be confounded in the 1st, 2nd .... rth replication of the columns by $C_1, C_2, \ldots . C_r$ respectively, then there exists a set of effects $S_1, S_2, \ldots . S_r$ each with $(r - 1)$ degrees of freedom which together with the effects $C_1, C_2, \ldots . C_r$ and corresponding interactions which we denote by C_iS_j, i, j=1,2 ....r are all orthogonal to the effects confounded in the rows.

(b) The r^2 groups obtained from r complete replications confounding the effects $S_1, S_2, \dots S_r$ are such that r complete replications can be obtained by choosing one group from each of the above r replications.

TABLE 3. QUASI-LATIN SQUARE DESIGNS FOR $2 \times 2 \times 2$

number of latin squares	effects confounded with			relative information
	rows	groups	columns	
1	AB ; BC	B	ABC ; ABC	AB = AC = $\frac{1}{2}$, ABC = 0
3	AB ; BC BC ; CA CA ; AB	B C A	ABC ; ABC ABC ; ABC ABC ; ABC	$\left.\begin{matrix}\text{AB}\\ \text{BC}\\ \text{CA}\end{matrix}\right\} = \frac{2}{3}$, ABC = 0
1	BC ; CA	C	AB ; ABC	AB = BC = CA = ABC = $\frac{1}{2}$
3	AB ; BC BC ; CA CA ; AB	B C A	AC ; ABC AB ; ABC BC ; ABC	$\left.\begin{matrix}\text{AB}\\ \text{BC}\\ \text{CA}\end{matrix}\right\} = \frac{2}{3}$, ABC = $\frac{1}{2}$
3	A ; AB B ; BC C ; AC	A or B B or C C or A	ABC ; ABC ABC ; ABC ABC ; ABC	$\left.\begin{matrix}\text{A} = \text{AB}\\ \text{B} = \text{BC}\\ \text{C} = \text{CA}\end{matrix}\right\} = \frac{1}{2}$, ABC = 0

In the case $s = 3$, $a = 3$, $b = 2$, with a latin square of size 9×9 only some degrees of freedom belonging to the triple factor interaction need be confounded. The 9 degrees of freedom belonging to the triple factor interaction can be split into 4 orthogonal sets, T_1, T_2, T_3 and T_4 of 2 degrees of freedom each. They are given below :

TABLE 4. ORTHOGONAL SET FOR THE 2ND ORDER INTERACTION $3 \times 3 \times 3$

combination of 1st. and 2nd factors		T_1	T_2	T_3	T_4
		levels of the third factor			
1	00	0 2 1	0 1 2	0 2 1	0 1 2
2	10	1 0 2	2 0 1	1 0 2	2 0 1
3	20	2 1 0	1 2 0	2 1 0	1 2 0
4	01	2 1 0	1 2 0	1 0 2	2 0 1
5	11	0 2 1	0 1 2	2 1 0	1 2 0
6	21	1 0 2	2 0 1	0 2 1	0 1 2
7	02	1 0 2	2 0 1	2 1 0	1 2 0
8	12	2 1 0	1 2 0	0 2 1	0 1 2
9	22	0 2 1	0 1 2	1 0 2	2 0 1

(a) Two sets T_i and T_j may be chosen for confounding in all the replications of the rows and columns respectively so that no information on T_i and T_j can be had from one square. The design is constructed with the help of choosing any third set T_k, $k \neq i$ or j or any suitable second order interaction for the effects S_1, S_2 and S_3 defined in para. (b) above. With two squares confounding T_1 and T_2 in one and T_3 and T_4 in another we have a balanced arrangement which gives $\frac{1}{2}$ the relative information on all the four.

(b) A second system of arrangement is to confound the sets T_1, T_2 and T_2 in the replications of the rows and the sets T_3, T_4 and T_4 in those of the columns. The relative information on T_2 and T_4 is $\frac{1}{2}$ and that on T_1 and T_3 is $\frac{2}{3}$. The construction is effected by choosing a second order interaction for S_1, S_2 and S_3. The solution is given below.

TABLE 5. THE QUASI-LATIN SQUARE DESIGN FOR $3\times3\times3$

10	61	82	50	71	32	90	21	42	T_1
60	81	12	70	31	52	20	41	92	
80	11	62	30	51	72	40	91	22	
31	72	50	22	90	41	81	62	10	T_2
51	32	70	42	20	91	11	82	60	
71	52	30	92	40	21	61	12	80	
22	90	41	81	62	10	72	50	31	T_2
42	20	91	11	82	60	32	70	51	
92	40	21	61	12	80	52	30	71	
T_1			T_3			T_4			

The confounded effect in the group of any replication row or column wise is the comparison defined by the following table.

TABLE 6. THE COMMON CONFOUNDED EFFECTS IN GROUPS OF TABLE 5.

levels of the third factor	comparison
0 + 1 + 2	1 3 2
0 + 1 + 2	6 5 4
0 + 1 + 2	8 7 9

The design of Table 5 is useful for layout only after proper randomization. We may choose another square confounding T_2, T_1, T_1 and T_3, T_3, T_4 together with the above for balancing. The relative information is $\frac{1}{2}$ on all.

TABLE 7. 8×8 QUASI-LATIN SQUARE FOR 2^4 DESIGN.

ABCD		ABCD		ABCD		ABCD		
44	42	23	31	33	21	14	12	ABC
41	43	22	34	32	24	11	13	
22	34	44	12	41	13	23	31	BCD
33	24	11	21	14	43	32	42	
14	13	32	43	22	31	44	21	ABD
11	12	41	24	23	42	33	34	
23	31	14	42	11	34	22	43	ACD
32	21	33	13	44	12	41	24	

In the case $r = 4, s = 2, a = 4, b = 3$, there are four factors A, B, C and D. The effects ABC, BCD and CDA can be confounded with the rows and the effect ABCD with the columns. The unrandomized design is given in Table 7 where the first and second figures stand for the following. $1 = (1)$, $2 = a$, $3 = b$, $4 = ab$ and $1 = (1)$, $2 = c$, $3 = d$, $4 = cd$. Thus 44 stands for *abcd*.

4. EXTENSION TO THE GENERAL FACTORIAL DESIGN.

We can, now, extend some of the principles enunciated above to the asymmetrical factorial designs of the form $p \times q^2$ in pq plots blocks. When q is a prime or a prime power it is possible to divide the $(q-1)^2$ degrees of freedom belonging to BC into $(q+1)$ orthogonal sets of $(q-1)$ degrees of freedom each arising from the comparisons.

$$Q^i_1, Q^i_2, \ldots . Q^i_q;\ i = 1, 2, \ldots . (q-1) \qquad .. \quad (4)$$

It has been shown by Nair and Rao (1941) that a design for pq^2 in pq plots blocks can be found in $(q-1)$ replications if $p \leqslant q$ obtaining balance on any of the comparisons (4) and its interaction with A. For complete balance $(q-1)^2$ replications are necessary. Since $k^2/r = pq^2$, $k = pq$, $r = p$; $(q-1)$ replications can be arranged in the columns and another set of $(q-1)$ replications giving balance over an orthogonal set can be arranged in the rows if $p = (q-1)$. It is easily seen that A is the only possible effect that can be confounded among the groups. Thus only one type of square is possible for any sets of replications. A completely balanced arrangement can be found in $(q-1)/2$ (if q is odd) or $(q-1)$ (if q is even) latin squares.

For the asymmetrical designs $p \times q$ in blocks of p plots the combinatorial restrictions give,

$$k^2/r = pq,\ k = p,\ p = rq \qquad .. \quad (4.1)$$

so that $pq = rq^2$ and the design with two factors A and B can be reduced to a design with three factors by the introduction of pseudovariates A_1 and A_2 at levels r and q, building up the levels of A. The construction of designs may be followed as in para. 1 of Section 3. The analysis may also be carried with the use of pseudovariates. Thus 6×3 design is possible in a quasi-latin square 6×6 confounding some effects of the interaction AB with the rows and some with the columns. The main effects are completely preserved. The design and analysis will be identical with that of 2×3^2. The designs 9×3, 8×4, etc. can always be arranged in 9×9, 8×8 etc. quasi-latin squares confounding the first order interaction only, with the use of pseudovariates.

The use of quasi-latin squares in confounded arrangements for factorial designs is explained in this article. The necessary randomization, the proper analysis together with the method of construction of the possible designs are discussed. The possibilities of arranging varietal trials in quasi-latin squares and the various types of designs have been already found (Rao : 1943) and a description of the methods is given below.

5. QASI-LATIN SQUARES IN VARIETAL TRIALS.

The methods developed in previous Sections can be extended to the case of experiments involving a number of varieties. These designs not only enhance the precision of comparisons but the number of replications is sometimes halved. The fundamental problem of arrangements in quasi-latin squares connected with varietal trial may be stated as follows.

There are v treatments which are denoted by $T_1, T_2, \ldots . T_v$, to be tested in n lattice squares of k^2 cells each such that the i-th variety is replicated r_i times on the total and the treatments T_i and T_j appear together λ_{ij} times in the rows and columns combined in all the squares and μ_{ij} times in all the squares.

$$\lambda_{ii}=0,\ \lambda_{ij}=\lambda_{ji},\ \mu_{ij}=\mu_{ji} \quad .. \quad (5{\cdot}1)$$

$$r_1+r_2+\ldots,+r_k=n\,k^2. \quad .. \quad (5.2)$$

$$2(k-1)\,r_i=\lambda_{i1}+\ldots+\lambda_{iv}$$

$$(k^2-1)r_i=\mu_{i1}+\ldots+\mu_{iv} \quad .. \quad (5.3)$$

The analysis of the most general type of design can be carried out as indicated below. The normal equations giving varietal comparisons (the varietal effects are denoted by $t_1, \ldots t_v$ for comparative purposes) are

$$Q_i=\frac{r_i\,(k-1)^2}{k^2}\,t_i-\frac{1}{k}\sum_j \lambda_{ij}\,t_j+\frac{1}{k^2}\sum_j \mu_{ij}\,t_j \quad .. \quad (5.4)$$

where Q_i = the sum of observations on the i-th treatment
— the sum of row and column means in which the i-th treatment occurs
+ the weighted total of square means, the weight being the number of times the i-th treatment occurs in a square.

If $t_1, t_2, \ldots t_v$ constitute a solution of the system of normal equations(5.4) then the analysis of variance table leading to tests of significance can be given as follows.

TABLE. 8. ANALYSIS OF VARIANCE TABLE

variation due to	d.f.	sum of squares
squares	$n-1$	U
columns	$n(k-1)$	U
rows	$n(k-1)$	U
treatments	$v-1$	$\Sigma t_i Q_i$
error	—	—
total	nk^2-1	U

The expressions marked by U are to be calculated in the usual manner. The sum of squares due to columns and rows are to be calcubated from 'within squares'. The expressions marked by — are obtained by subtraction.

The difference of i-th and j-th treatment effects is estimated by (t_i-t_j) and if this is equivalent to $l_1Q_1+\ldots+l_vQ_v$ then the variance of (t_i-t_j) is given by $(l_i-l_j)\,\sigma^2$, where σ^2 is the error variance so that using the estimate of σ^2 given by the error line in the above analysis of variance table we can test for treatment comparisons.

The simplest case of the quasi-latin square arrangements for varietal trials is when all the varieties occur in each suqare and $\lambda_{ij}=1$ for all i and j. These designs are derivable from the solution of the balanced incomplete block designs for the system of parameters

$$v=m^2,\ b=m^2+m,\ k=m,\ r=m+1 \quad .. \quad (5{\cdot}5)$$

where $m=p^s$ (p being a prime and s an integer). This is a resolvable series so that the design can be reduced to $(m+1)$ groups of m blocks each containing m plots. If $(m+1)$ is even we can combine the groups two by two to form $n=(m+1)/2$ quasi-latin squares. This is possible due to the orthogonal property of these groups, the rows and columns of a quasi-latin square representing two separate replications in the original design. Thus the number of replications can be halved.

In this special case $\mu_{ij}=1$ for all i and j so that the normal equations become

$$Q_i=\left(\frac{n(k-1)^2}{k^2}+\frac{1}{k}\right)t_i-\frac{1}{k}\,\Sigma t_j+\frac{1}{k^2}\,\Sigma t_j \quad \ldots \quad (5{\cdot}6)$$

so that

$$t_i-t_j=(Q_i-Q_j)\frac{k^2}{n(k-1)^2+k}$$

and

$$V(t_i-t_j)=\frac{2k^2}{n(k-1)^2+k}\,\sigma^2 \quad \ldots \quad (5{\cdot}7)$$

Various other types of designs can be derived by allowing λ's and μ's to assume assigned values.

References.

1. Bose, R. C. and Kishen, K. (1941) On the problem of confounding in the general symmetrical factorial design *Sankhya*, **5**(1), 21-36.
2. Nair, K. R. (1938) On a method of getting confounded arrangements in the general symmetrical type of experiments. *Sankhya*, **4**(1), 121-138.
3. Nair, K. R. and Rao, C. R. (1941) Confounded designs for asymmetrical factorial experiments. *Science & Culture*, **7**, 313-314.
4. Rao, C. R. (1943) Quasi-latin squares in experimental arrangements. *Current Science*, 12(12), 322-323.
5. Yates, F. (1937) The design and analysis of factorial experiments. Tech. Comm. No. 35. Imperial Bureau of Social Science.

Paper received 7 December, 1942.

ENUMERATION OF NON-ISOMORPHIC SOLUTIONS OF BALANCED INCOMPLETE BLOCK DESIGNS.

By HARI KINKAR NANDI
Statistical Laboratory, Calcutta.

INTRODUCTION

Two solutions for a balanced incomplete block design with parameters v, b, r, k, λ are said to be isomorphic if they are identical except for the renaming of blocks and varieties. The problem of the enumeration of all possible non-isomorphic solutions has been discussed by Fisher (1940), Cox (1940) and recently by Q. M. Husain (1945) who has tried his method on certain symmetrical designs only (i.e. $v = b$, $r = k$, $\lambda = 2$). A variant of that method having general application has been applied in this paper to the unsymmetrical designs : (A) $v = 6$, $b = 10$, $r = 5$, $k = 3$, $\lambda = 2$, and (B) $v = 10$, $b = 15$, $r = 6$, $k = 4$, $\lambda = 2$, which are in fact the designs obtained by block section from those considered by Husain.

In the symmetrical design, any two blocks have just λ treatments in common which helps in the building up of a design step by step. In the unsymmetrical cases, we require the frequency-distribution of the number of treatments common to a particular block (called the initial block) and the other $(b-1)$ blocks. A general solution has been given in section 2 of the paper from which it is seen that the frequency distribution is unique for designs with $\lambda = 1$, and also for those obtained by block-section from symmetrical designs with $\lambda = 2$.

The actual enumeration is given in section 3 ; and it has been found that (A) has only one solution with which the solutions given by Fisher and Yates (Tables, 1938), and Bose (1939) are isomorphic, while (B) has three non-isomorphic solutions given in this paper as $[\alpha_{11}]$, $[\alpha_3]$, $[\beta_{21}]$; the solution given by Fisher and Yates (Tables, 1938) is isomorphic with $[\alpha_3]$ while that given by Bose (1939) is isomorphic with $[\beta_{21}]$.

Bose (1939) has shown how from a symmetrical design with $v = b$, $r = k$, λ we can always obtain by block-section the designs with parameters $v' = v - k$, $b' = b - 1$, $r' = r$, $k' = k - \lambda$, $\lambda' = \lambda$. It remains to be seen how far the reverse process of block-section i.e. adjunction of the blocks of the designs $v'' = k$, $b'' = b - 1$, $r'' = r - 1$, $k'' = \lambda$, $\lambda'' = \lambda - 1$ to the latter design, is in general possible, giving the symmetrical design. In section 4 it has been shown that adjunction is possible in both the cases considered above, and a general method is given for this purpose. (A) after adjunction leads to only one solution of $v = b = 11$, $r = k = 5$, $\lambda = 2$, while (B) leads to three non-isomorphic solutions thus verifying the conclusion of Hussain (1943) regarding the two symmetrical designs : $v = b = 11$, $r = k = 5$, $\lambda = 2$ and $v = b = 16$, $r = k = 6$, $\lambda = 2$.

2. FREQUENCY DISTRIBUTION OF THE NUMBER OF COMMON TREATMENTS

Let us consider a variable x giving the number of treatments common to the initial block and any other block of the design. Then there are $(b-1)$ positive integral values of this variable. Let f_i stand for the frequency of the value $x = i$ ($i = 0, 1, 2, \ldots, n$, $n \leq k$).

Then we must have,

$$\left.\begin{aligned} f_0 + f_1 + f_2 + \dots + f_n &= b-1 \\ f_1 + 2f_2 + \dots + nf_n &= k(r-1) \\ f_1 + 2^2 f_2 + \dots + n^2 f_n &= k(r-1) + k(k-1)(\lambda-1) \end{aligned}\right\} \quad .. \quad (1)$$

All possible solutions satisfying (1) can be obtained by adding to any particular solution, the solution,

$$\left.\begin{aligned} f_0 &= -f_3 - 3f_4 - \dots - (n^2-3n+2)/2.\, f_n \\ f_1 &= \quad 3f_3 + 8f_4 + \dots + (n^2-2n)\, f_n \\ f_2 &= -3f_3 - 6f_4 - \dots - (n^2-n)/2.\, f_n \text{ etc.} \\ & f_3, f_4, \dots, f_n \end{aligned}\right\} \quad .. \quad (2)$$

$f_3, f_4, \dots, f_n$ being arbitrary. The admissible solutions are those giving positive integral values for $f_0, f_1, \dots, f_n$. For the particular solution we may try,

$$\left.\begin{aligned} f_0 + f_1 + f_2 &= b-1 \\ f_1 + 2f_2 &= k(r-1) \\ f_1 + 4f_2 &= k(r-1) + k(k-1)(\lambda-1) \end{aligned}\right\} \quad .. \quad (3)$$

Now we shall consider certain special cases which will be of use later.

Case I. $\lambda = 1$. For the particular solution we have from (3), $f_0 = (b-1) - k(r-1)$, $f_1 = k(r-1)$, $f_2 = 0$ the values being positive integral ; and from (2) it is seen that this is the unique solution as otherwise f_2 will be negative.

Case II. $\lambda = 2$ and the class of designs : $v' = v-k$, $b' = b-1$, $r' = r$, $k' = k-\lambda$, $\lambda' = \lambda$ obtained by block-section from $v = b$, $r = k$, λ. For the particular solution we have

$$f_0 = 0, \quad f_1 = 2(k-2) = 2(r'-2), \quad f_2 = (k-2)(k-3)/2 = (r'-2)(r'-3)/2$$

and here also the solution is unique otherwise f_0 will be negative.

It should be noted that this uniqueness of solution is not a general property as is easily seen from the designs : (a) $v = 10$, $b = 30$, $r = 9$, $k = 3$, $\lambda = 2$; and (b) $v = 9$, $b = 18$, $r = 8$, $k = 4$, $\lambda = 3$ each of which has two solutions for frequency distribution namely (a) $f_0 = 8$, $f_1 = 18$, $f_2 = 3$, $f_3 = 0$ and $f_0 = 7$, $f_1 = 21$, $f_2 = 0$, $f_3 = 1$, and (b) $f_0 = 1$, $f_1 = 4$, $f_2 = 12$, $f_3 = 0$ and $f_0 = 0$, $f_1 = 7$, $f_2 = 9$, $f_3 = 1$. The solution given in Fisher and Yates' Table (1938) for the latter design shows both the distributions of 'x'.

3. ENUMERATION

We shall now consider the enumeration of the solutions of the design : $v = 6$, $b = 10$, $r = 5$, $k = 3$, $\lambda = 2$. Suppose any block to be filled up by the treatments 1, 2, 3. In the other nine blocks these treatments are to be placed in such a way that each is replicated four times and each of the three pairs occurs once only. From the results in section 2 it is seen that three of the other nine blocks have just two treatments in common with the initial block and the remaining six have just one treatment in common. Hence the nine blocks can be uniquely filled up as shown below :

(1, 2, 3) ; (1, 2) ; (1, 3) ; (2, 3) ; (1) ; (1) ; (2) ; (2) ; (3) ; (3).

Let the second block be filled up by the new treatment 4. In order to place 4 in the other blocks suppose y blocks be chosen from those containing two treatments (called two-treatment blocks) and x from those containing one treatment (called one-treatment blocks), then $x+y=4$, $x+2y=4$. which give $x=4$, $y=0$, thus yielding the partially filled up design :

(1, 2, 3) ; (1, 2, 4) ; (1, 3) ; (2, 3) ; (1, 4) ; (1) ; (2, 4) ; (2) ; (3, 4) ; (3, 4).

Fill up the third block with the new treatment 5 and then again for 5, $x+y=4, x+2y=6$. giving $x=2$, $y=2$. Since there are only two one-treatment blocks they are chosen. To choose the other two two-treatment blocks we will proceed in the following way which will be useful in the enumeration of the second design.

Let us write all the treatments except the new treatment introduced at this stage. Thus put one or two dashes : $1''$, $2'$, $3'$, 4 on them according as up to this stage of placing 5 it occurs once or twice with the corresponding treatment. This is shown above where it will be found that 4 remains undashed. Hence in choosing the two remaining blocks we must choose two each containing 4, and 2 or 3 (i.e. the blocks (3,4) (2,4). Hence placing the remaining treatment '6' to fill up the block, the unique solution is :

(1, 2, 3) ; (1, 2, 4) ; (1, 3, 5) ; (2, 3 ,6) ; (1, 4, 6) ; (1, 5, 6) ; (2, 4, 5) ; (2, 5, 6) ; (3, 4, 5) ; (3, 4, 6).

We next consider the enumeration of the solution of the design : $v=10$, $b=15$, $r=6$, $k=4$, $\lambda=2$. Fill up the initial blocks with the treatments 1, 2, 3, 4. Of the remaining fourteen blocks, eight blocks have one treatment in common with the initial block and the other six have two treatments in common. So the blocks filled with these four treatments will be as follows :

(1, 2, 3, 4) ; (1, 2) ; (1, 3) ; (1, 4) ; (2, 3) ; (2, 4) ; (3, 4) ; (1) ; (1) ; (2) ; (2) ; (3) ; (3) ; (4) ; (4).

Fill up the second block with the new treatments 5, 6. The number of one-treatment blocks (x) and two-treatment blocks (y) for placing of 5 is given by $x+y=5$, $x+2y=6$. which gives $x=4$, $y=1$. Any one of the two-treatment blocks (1,3), (1,4), (2,3), (2,4), (3,4) can be chosen. But choices of one of (1,3), (1,4), (2,3), (2,4) are isomorphic as can be seen by interchanging the names 3, 4, or 1, 2. Consequently we have two non-isomorphic ways of choosing one two-treatment block and the four one-treatment block are then chosen uniquely giving

[α] : (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 5) ; (1, 4) ; (2, 3); (2, 4) ; (3, 4); (1) ; (1); (2, 5) ; (2); (3, 5);
(3) ; (4, 5) ; (4, 5).

[β] : (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3) ; (1, 4) ; (2, 3) ; (2, 4) ; (3, 4, 5) ; (1, 5) ; (1) ; (2, 5) ; (2) ;
(3, 5) ; (3) ; (4, 5) ; (4).

At this stage the number of one-treatment blocks (x), two-treatment blocks (y), and three-treatment blocks (z) to which 6 is to be alloted is given by $x+y+z=5$, $x+2y+3z=7$ The admissible solutions are (i) $z=0$, $y=2$, $x=3$; (ii) $z=1$, $y=0$, $x=4$.

Case [α]. Here solution (ii) is obviously impossible and as such the only way of choosing according to (i) will be the blocks (1), (2), (3) (3,4), (4,5) giving us the design filled upto '6' as :

(1, 2, 3, 4) : (1, 2, 5, 6) ; (1, 3, 5) ; (1, 4) ; (2, 3) ; (2, 4) ; (3, 4, 6) ; (1, 6) ; (1) ; (2, 5) ;
(2, 6) ; (3, 5) ; (3, 6) ; (4, 5, 6) ; (4. 5).

Completing the third block by 7, we see that it can be placed only in one way in the other blocks corresponding to one admissible solution of $x+y+z=5$, $x+2y+3z=9$ giving us.

(1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 5, 7) ; (1, 4) ; (2, 3) ; (2, 4, 7) ; (3, 4, 6) ; (1, 6) ;
(1, 7) ; (2, 5) ; (2, 6, 7) ; (3, 5) ; (3, 6, 7) ; (4, 5, 6) ; (4, 5, 7).

The fourth block is now completed with 8, 9 and 8 is to be placed in two three-treatment blocks and three two-treatment blocks. The possible ways of choosing these are :

$[\alpha_1]$ (1, 6) (2, 3) (2, 5) ; $[\alpha_2]$ (1, 5) (2, 3) (3, 5) ; $[\alpha_3]$ (1, 6) (2, 5) (3, 5) ; $[\alpha_4]$ (1, 7)
(2, 3) (2, 5) ; $[\alpha_5]$ (1, 7) (2, 3) (3, 5) ; $[\alpha_6]$ (1, 7) (2, 5) (3, 5) ; $[\alpha_7]$ (2, 3) (2; 5) (3, 5)

Now $[\alpha_7]$ can be shown to be impossible while $[\alpha_4]$ is isomorphic with $[\alpha_2]$, $[\alpha_5]$ with $[\alpha_1]$ and $[\alpha_6]$ with $[\alpha_3]$ as the transformation $\begin{pmatrix} 1\,2\,3\,4\,5\,6\,7 \\ 1\,3\,2\,4\,5\,7\,6 \end{pmatrix}$ shows. Also $[\alpha_3]$ is isomorphic with $[\alpha_2]$ as the transformation $\begin{pmatrix} 1\,2\,3\,4\,5\,6\,7 \\ 1\,2\,5\,6\,3\,4\,7 \end{pmatrix}$ reveals. Hence the two non-isomorphic ways of placing '8' give

$[\alpha_1]$ (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 5, 7) ; (1, 4, 8, 9) ; (2, 3, 8) ; (2, 4, 7) ; (3, 4, 6) ;
(1, 6, 8) ; (1, 7) ; (2, 5, 8) ; (2, 6, 7) ; (3, 5) ; (3, 6, 7, 8) ; (4, 5, 6) ; (4, 5, 7, 8)

$[\alpha_3]$ (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 5, 7) ; (1, 4, 8, 9) ; (2, 3) ; (2, 4, 7, 8) ; (3, 4, 6);
(1, 6, 8) ; (1, 7) ; (2, 5, 8) ; (2, 6, 7) ; (3, 5, 8) ; (3, 6, 7, 8) ; (4, 5, 6) ; (4, 5, 7).

In placing 9 in $[\alpha]$ we have to choose two two-treatment blocks and three three-treatment blocks, and as there are only two two-treatment blocks, they will be filled up. Then the three three-treatment blocks can be chosen in two non-isomorphic ways namely $[\alpha_{11}]$ (2,3,8).(2,6,7), (4,5,6) and $[\alpha_{12}]$ (2,5,8) (3,4,6) (2,6,7) . Of these, the placing of 9 in $[\alpha_{12}]$ is isomorphic with that of 8 in $[\alpha_2]$ or $[\alpha_3]$. Again $[\alpha_3]$ admits of only one placing of 9. Hence from $[\alpha]$ two non-isomorphic solutions are obtained :

$[\alpha_{11}]$ (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 5, 7) ; (1, 4, 8, 9) ; (2, 3, 8, 9) ; (2, 4, 7, 10) ;
(3, 4, 6, 10) ; (1, 6, 8, 10) ; (1, 7, 9, 10) ; (2, 5, 8, 10) ; (2, 6, 7, 9) ; (3, 5, 9, 10)
(3, 6, 7, 8) ; (4, 5, 6, 9) ; (4, 5, 7, 8).

$[\alpha_3]$ (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 5, 7) ; (1, 4, 8, 9) ; (2, 3, 9, 10) ; (2, 4, 7, 8) ;
(3, 4, 6, 10) ; (1, 6, 8, 10) ; (1, 7, 9, 10) ; (2, 5, 8, 10) ; (2, 6, 7, 9) ; (3, 5, 8, 9)
(3, 6, 7, 8) ; (4, 5, 6, 9) ; (4, 5, 7, 10).

Case $[\beta]$. Here for placing of 6 solution (*i*) will give its placing isomorphic with that of 5 in $[\alpha]$ and according to [*ii*] placing of 6 will yield

(1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3) ; (1, 4) ; (2, 3) ; (2, 4) ; (3, 4, 5, 6) ; (1, 5) ; (1, 6) ;
(2, 5) ; (2, 6) ; (3, 5) ; (3, 6) ; (4, 5) ; (4, 6) .

Filling up the third block with 7 and 8 other five blocks which all contain two treatments to accomodate 7 can be chosen in the following way. We have to choose one two-treatment block containing 1 i.e. we have to choose either (14) or (15) or (16). If we choose (14) placing of 7 will be isomorphic with that of 5 in $[\alpha]$ while choice of (15) or (16) are isomorphic as is easily seen by interchanging 5, 6. So there are only three non-isomorphic ways of choosing the five blocks :

$[\beta_1]$ (1, 5) (2, 4) (2, 5) (3, 6) (4, 6) $[\beta_2]$ (1, 5) (2, 4) (2, 6) (3, 6) (4, 5)
$[\beta_3]$ (1, 5) (2, 4) (2, 6) (3, 5) (4, 6).

leading to the partially filled-up designs :

$[\beta_1]$ (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 7, 8) ; (1, 4) ; (2, 3) ; (2, 4, 7) ; (3, 4, 5, 6) ;
(1, 5, 7) ; (1, 6) ; (2, 5, 7) ; (2, 6) ; (3, 5) ; (3, 6, 7) ; (4, 5) ; (4, 6, 7).

[β_2] (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 7, 8) ; (1, 4) ; (1, 3) ; (2, 4, 7) ; (3, 4, 5, 6) ;
(1, 5, 7) ; (1, 6) ; (2, 5) ; (2, 6, 7) ; (3, 5) ; (3, 6, 7) ; (4, 5, 7) ; (4, 6).

[β_3] (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 7, 8) ; (1, 4) ; (2, 3) ; (2, 4, 7) ; (3, 4, 5, 6) ;
(1, 5, 7) ; (1, 6) ; (2, 5) ; (2, 6, 7) ; (3, 5, 7) ; (3, 6) ; (4, 5) ; (4, 6, 7).

As regards placing of 8, we have to choose one three-treatment block ; and it may be noted that except the choice of the block (2, 4, 7,) all other choices make the placing of 8 isomorphic with that of 5 in [α]. Hence 8 is placed uniquely in the above three solutions. For placing of 9 each of the three admits of two non-isomorphic ways, but again in case of [β_2] the second way of placing 9 is isomorphic with that of 7 in [β_1] ; and in the case of [β_3] the second way is isomorphic with that of 7 in [β_2]. Thus we obtain four non-isomorphic solutions ; two from [β_1] and one each from [β_2] and [β_3]

[β_{11}] (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 7, 8) ; (1, 4, 9, 10) ; (2, 3, 9, 10) ; (2, 4, 7, 8) ;
(3, 4, 5, 6) ; (1, 5, 7, 9) ; (1, 6, 8, 10) ; (2, 5, 7, 10) ; (2, 6, 8, 9) ; (3, 5, 8, 10) ;
(3, 6, 7, 9) ; (4, 5, 8, 9) ; (4, 6, 7, 10) ;

[β_{12}] (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 7, 8) ; (1, 4, 9, 10) ; (2, 3, 9, 10) ; (2, 4, 7, 8) ;
(3, 4, 5, 6) ; (1, 5, 7, 9) ; (1, 6, 8, 10) ; (2, 5, 7, 10) ; (2, 6, 8, 9) ; (3, 5, 8, 9) ;
(3, 6, 7, 10) ; (4, 5, 8, 10) ; (4, 6, 7, 9).

[β_{21}] (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 7, 8) ; (1, 4, 9, 10) ; (2, 3, 9, 10) ; (2, 4, 7, 8) ;
(3, 4, 5, 6) ; (1, 5, 7, 9) ; (1, 6, 8, 10) ; (2, 5, 8, 10) ; (2, 6, 7, 9) ; (3, 5, 8, 9) ;
(3, 6, 7, 10) ; (4, 5, 7, 10) ; (4, 6, 8, 9).

[β_{31}] (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 3, 7, 8) ; (1, 4, 9, 10) ; (2, 3, 9, 10) ; (2, 4, 7, 8) ;
(3, 4, 5, 6) ; (1, 5, 7, 9) ; (1, 6, 8, 10) ; (2, 5, 8, 10) ; (2, 6, 7, 9) ; (3, 5, 7, 10) ;
(3, 6, 8, 9) ; (4, 5, 8, 9) ; (4, 6, 7, 10).

Now it remains to be seen whether choice of any other block of the above solutions as the initial makes them isomorphic with the [α] set. Taking the last block of each solution as the initial block, it is revealed that [β_{11}] is isomorphic with [α_3] while [β_{12}], [β_{31}] are isomorphic with [α_{11}]. The following transformations establish this isomorphism.

$$[\beta_{11}] \qquad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & & 7 & 8 & 9 & 10 \\ 8 & 10 & 5 & 2 & 6 & 1 & 3 & 9 & 7 & 4 \end{pmatrix}$$

$$[\beta_{12}] \qquad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 10 & 9 & 5 & 3 & 7 & 1 & 2 & 8 & 4 & \end{pmatrix}$$

$$[\beta_{31}] \qquad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 10 & 5 & 9 & 3 & 8 & 2 & 1 & 7 & 6 & 4 \end{pmatrix}$$

Hence, we get three solutions in all, for the design ; $v=10$, $'=15$, $r=6$, $k=4$, $\lambda=2$ given in the paper as [α_{11}], [α_3], [β_{21}].

4. PROCESS OF BLOCK ADJUNCTION

Husain (1945) has shown that the symmetrical design, $v=b=11$, $r=k=5$, $\lambda=2$ has only one solution while $v=b=16$, $r=k=6$, $\lambda=2$ has three non-isomorphic solutions. This, along with the conclusions arrived at above regarding the designs obtained by block-section, raises the problem whether it is always possible to pass from the unsymmetrical to the corresponding symmetrical design by suitable adjunction of blocks ; and if possible how do the non-isomorphic solutions of the two stand in relation to one another.

It is obvious that the designs whose blocks need to be adjoined is the design obtained from the symmetrical design by block-intersection. So the necessary conditions for the reverse process of block-section are : (1) the design of block-intersection must exist, and (2) the distribution of x, the number of treatments common to the initial block and any other block in the two designs to be adjoined must be complementary in the sense that if in the former $f_i=l$, then in the later $f_{\lambda-i}$ must also be equal to l. If adjunction is possible, there may arise two situations ; (*i*) more than one non-isomorphic ways of adjunction may be possible ; (*ii*) after adjunction two non-isomorphic solutions may become isomorphic, so that the number of solutions in the unsymmetrical case may be less than, equal to or greater than the number of solutions in the symmetrical case.

It is easy to see that the necessary conditions enumerated above are satisfied in the case of the designs (A) and (B) considered in this paper. The former requires for adjunction ten blocks of all possible pairs of five new treatments and one block containing these five treatments to get the design ; $v=b=11$, $r=k=5$, $\lambda=2$, while the later, all possible pairs of six new treatments to obtain ; $v=b=16$, $r=k=6$, $\lambda=2$. For actual adjunction the following method may be attempted with slight variation in complex cases and if it succeeds, adjunction is possible ; otherwise not.

(A) The solution is $[\alpha]$:

(1, 2, 3) (7, 8]
(1, 2, 4) [9, 10]
(1, 3, 5) [9, 11]
(2, 3, 6) (10, 11]
(1, 4, 6) [7, 11]
(1, 5, 6) [8. 10]
(2, 4, 5) [8, 11]
(2, 5, 6) [7, 9]
(3, 4, 5) [7, 10]
(3, 4, 6) [8, 9]
[7, 8, 9, 10, 11]

To the first block we adjoin the treatment pair (7, 8). The second having two treatments in common with the first, two other new varieties (9, 10) are adjoined to it. The third block has two treatments in common with the first but only one with the second. Hence the treatment pair to be adjoined to the third will contain the new treatment 11 and one of 9, 10. In this way treatment-pairs are allotted to the remaining blocks keeping in mind that any two new blocks have just two treatments in common. In the above, the adjoined pairs are shown in [ ] and it will be found that the conditions of balanced incomplete block designs are satisfied.

(B) The solutions are $[\alpha_{11}]$, $[\alpha_3]$, $[\beta_{21}]$. Here, in the first place, a new treatment pair is adjoined to the first block. Then out of the six other blocks which have two treatments in common with the first, two blocks are chosen which have also two treatments in common with each other, to which new treatment-pairs are assigned . In all the three cases this is possible ; and then following the above method as in the case of (A) we have only one way of adjoining to each of the three :

$[\alpha'_{11}]$
(1, 2, 3, 4) [11, 12]
(1, 2, 5, 6) [13, 15]
(1, 3, 5, 7) [14, 16]
(1, 4, 8, 9) [13, 14]
(2, 3, 8, 9) [15, 16]
(2, 4, 7, 10) [13, 16]
(3, 4, 6, 10) [14, 15]
(1, 6, 8, 10) [11, 16]
(1, 7, 9, 10) [12, 15]
(2, 5, 8, 10) [12, 14]

$[\alpha'_3]$
(1, 2, 3, 4) [11, 12]
(1, 2, 5, 6) [13, 15]
(1, 3, 5, 7) [14, 16]
(1, 4, 8, 9) [13, 14]
(2, 3, 9, 10) [13, 16]
(2, 4, 7, 8) [15, 16]
(3, 4, 6, 10) [14, 15]
(1, 6, 8, 10) [11, 16]
(1, 7, 9, 10) [12, 15]
(1, 5, 8, 10) [12, 14]

$[\beta'_{21}]$
(1, 2, 3, 4) [11, 12]
(1, 2, 5, 6) [13, 14]
(1, 3, 7, 8) [13, 15]
(1, 4, 9, 10) [14, 15]
(2, 3, 9, 10) [13, 16]
(2, 4, 7, 8) [14, 16]
(3, 4, 5, 6) [15, 16]
(1, 5, 7, 9) [11, 16]
(1, 6, 8, 10) [12, 16]
(2, 5, 8, 10) [11, 15]

$[\alpha'_{11}]$	$[\alpha'_{3}]$	$[\beta'_{21}]$
(2, 6, 7, 9) [11, 14]	(2, 6, 7, 9) [11, 14]	(2, 6, 7, 9) [12, 15]
(3, 5, 9, 10) [11, 13]	(3, 5, 8, 9) [11, 15]	(3, 5, 8, 9) [12, 14]
(3, 6, 7, 8) [12, 13]	(3, 6, 7, 8) [12, 13]	(3, 6, 7, 10) [11, 14]
(4, 5, 6, 9) [12, 16]	(4, 5, 6, 9) [12, 16]	(4, 5, 7, 10) [12, 13]
(4, 5, 7, 8) [11, 15]	(4, 5, 7, 10) [11, 13]	(4, 6 8, 9) [11, 13]
[11, 12, 13, 14, 15, 16]	[11, 12, 13, 14, 15, 16]	[11, 12, 13, 14, 15, 16]

Let us now consider the following transpositions in the six cases :—

$[\alpha'_{11}]$: Interchange (11, 5) and (12, 6) and make the substitution

$$\begin{pmatrix} 7, & 8, & 9, & 10, & 11, & 12, & 13, & 14, & 15, & 16 \\ 13, & 14, & 15, & 16, & 8, & 10, & 9, & 12, & 7, & 11 \end{pmatrix}$$

$[\alpha'_{3}]$: Interchange (11, 6), (12, 5) and (2, 3) then make the substitution

$$\begin{pmatrix} 7, & 8, & 9, & 10, & 11, & 12, & 13, & 14, & 15, & 16 \\ 10, & 14, & 15, & 16, & 8, & 13, & 11, & 9, & 12, & 7 \end{pmatrix}$$

$[\beta'_{21}]$: Interchange (11, 5) and (12, 6) and then substitute

$$\begin{pmatrix} 7, & 8, & 9, & 10, & 11, & 12, & 13, & 14, & 15, & 16 \\ 12, & 13, & 14, & 15, & 9, & 10, & 7, & 8, & 11, & 16 \end{pmatrix}$$

After these transformations, it will be found that the symmetrical designs $[\alpha'_{11}]$, $[\alpha'_{3}]$, $[\beta'_{21}]$ are isomorphic respectively with the solutions (II), (III), (I) given by Husain (1945).

5. TEST OF ISOMORPHISM

It will be seen that when considering the placings and adjunctions, we reject the isomorphic ones, but that does not dispose of all the isomorphic solutions. So when the final solutions have been obtained we have to test their isomorphisms.

In the first place, test the isomorphisms of the broad types. Here we have two broad types [α] and [β]. We take the blocks of each solution of (β), one by one and correspond it to the initial block of type [α] and see whether (β) is isomorphic with [α]. This method has revealed, as shown above, that out of the four solutions from (β), three are isomorphic with [α]. Secondly, to point out the isomorphism of the two solutions from the same set we can adopt the following method.

Let the first solution remain as it is, and identify the blocks of the second, one by one, with the initial block of the first solution. Then consider the placings of other treatments in the second solution with reference to the treatments in the initial blocks. If this set of placings can be made to correspond to the set of placings in the first solution, the two are isomorphic; otherwise not.

As an illustration, consider the isomorphism of $[\beta_{12}]$ and $[\beta_{31}]$. The two initial blocks of the two solutions are made to correspond. Then the other six treatments occupy the following positions with reference to the four treatments in the initial block.

Treatments	Blocks of $[\beta_{12}]$	Blocks of $[\beta_{31}]$
5	(12), (34), $(1)_a$ $(2)_a$ $(3)_a$ $(4)_a$	(12), (34), $(1)_a$ $(2)_a$ $(3)_a$ $(4)_a$
6	(12), (34), $(1)_b$ $(2)_b$ $(3)_b$ $(4)_b$	(12), (34), $(1)_b$ $(2)_b$ $(3)_b$ $(4)_b$
7	(13), (24), $(1)_a$ $(2)_a$ $(3)_a$ $(4)_a$	(13), (24), $(1)_a$ $(2)_a$ $(3)_a$ $(4)_a$
8	(13), (24), $(1)_b$ $(2)_b$ $(3)_b$ $(4)_b$	(13), (24), $(1)_b$ $(2)_b$ $(3)_b$ $(4)_b$
9	(14), (23), $(1)_a$ $(2)_a$ $(3)_a$ $(4)_a$	(14), (23), $(1)_a$ $(2)_a$ $(3)_a$ $(4)_a$
10	(14), (23), $(1)_b$ $(2)_b$ $(3)_b$ $(4)_b$	(14), (23), $(1)_b$ $(2)_b$ $(3)_b$ $(4)_b$

The two blocks, each containing 1 or 2 or 3 or 4 are distinguished as 'a' or 'b' in the order of their positions. Let us concentrate on the patterns of the one-treatment blocks.

Treatments	$[\beta_{12}]$				$[\beta_{31}]$			
5	*a*	*a*	*a*	*a*	*a*	*a*	*a*	*a*
6	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*b*
7	*a*	*a*	*b*	*b*	*a*	*b*	*a*	*b*
8	*b*	*b*	*a*	*a*	*b*	*a*	*b*	*a*
9	*a*	*b*	*a*	*b*	*a*	*b*	*b*	*a*
10	*b*	*a*	*b*	*a*	*b*	*a*	*a*	*b*

If we interchange 1, 4, the sets of two treatments blocks correspond but the pattern in $[\beta_{31}]$ becomes

7	*b*	*b*	*a*	*a*
8	*a*	*a*	*b*	*b*
5	*a*	*a*	*a*	*a*
6	*b*	*b*	*b*	*b*
9	*a*	*b*	*a*	*b*
10	*b*	*a*	*b*	*a*

Interchange (a, b) in the third and fourth positions, and we see that the transformed pattern is indentical with $[\beta_{12}]$ and we get the following substitution :

$$\begin{pmatrix}1 & 4 & 5 & 6 & 7 & 8\\ 4 & 1 & 7 & 8 & 6 & 5\end{pmatrix}$$

to establish the isomorphism of $[\beta_{31}]$ with $[\beta_{12}]$. The isomorphism of the symmetrical designs are tested by considering the isomorphism of all possible block-cuts.

References

1. Bose, R. C. (1939) On the construction of balanced incomplete block design. *Annals of Eugenics*, 9, 358-399.
2. Cox, Gertrude M. (1940) Enumeration and construction of balanced incomplete configuration. *Annals of Math. St.*, 11, 72-85.
3. Fisher, R. A. (1940) An examination of the different possible solutions of a problem in incomplete block. *Annals of Eugenics*, 10, 52-75.
4. Fisher & Yates (1938) Statistical Tables for Biological, Agricultural & Medical Research. Olive & Boyd, London.
5. Husain, Q. M. (1945) On the totality of solutions of symmetrical incomplete block designs. *Sankhyā*, 7(2), 204-208.

Paper received : 14 *October*, 1943.

A FURTHER NOTE ON NON-ISOMORPHIC SOLUTIONS OF INCOMPLETE BLOCK DESIGNS

By HARIKINKAR. NANDI
Statistical Laboratory, Calcutta

INTRODUCTION

In a previous paper (Nandi 1944), a method has been given for enumerating all possible non-isomorphic solutions (and, of course, for the construction) of a 'balanced incomplete block' design. The method which may be termed the 'method of structural analysis' consists in filling up the blocks of a design with the desired number of treatments, step by step, keeping in mind the structure of the design or configuration. It has also been seen that given two unsymmetrical configurations with parameters

(i) $v'=v-k,\ b'=b-1,\ r'=r,\ k'=k-\lambda,\ \lambda'=\lambda$

and (ii) $v''=k,\ b''=b-1,\ r''=r-1,\ k''=\lambda,\ \lambda''=\lambda-1,$

when and how we can construct a symmetrical configuration with parameters $v=b,\ r=k,\ \lambda$. (The two unsymmetrical configurations are really the designs obtained by block-section and block-intersection (Bose 1939) respectively from the symmetrical design. The first one will be called 'residual configuration' and the second, the 'derived configuration' (Ayyanger 1943). This procedure of block-adjunction to enumerate the symmetrical configuration is a very convenient one inasmuchas the unsymmetrical configurations containing lesser number of treatments admit of easier enumeration.

Let us consider the symmetrical configuration : $v=b=15,\ r=k=7,\ \lambda=3$. The residual configuration has the parameters ; $v=8,\ b=14,\ r=7,\ k=4,\ \lambda=3$ whereas for the derived confiruation : $v=7,\ b=14,\ r=6,\ k=3,\ \lambda=2$ both of which are enumerated with less labour.

2. ENUMERATION OF THE RESIDUAL AND DERIVED CONFIGURATIONS

Consider first the configuration : $v=8,\ b=14,\ r=7,\ k=4,\ \lambda=3$. Let us call any block of this as the initial block and let 'x' be a variable giving the number of treatments common to the initial block and any other block. Then there are thirteen values of 'x' and denoting by f_i the frequency of $x=i$, it is easily seen (Nandi 1944) that the following two sets of values of f_i's only are admissible for the case under consideration, viz.,

(a) $f_0=0,\ f_1=3,\ f_2=9,\ f_3=1$

(b) $f_0=1,\ f_1=0,\ f_2=12,\ f_3=0$

Suppose we start with an initial block which follows the law (a). Then filling up the first block with treatments 1, 2, 3, 4 we get the unique pattern :

(1, 2, 3, 4) ; (1, 2, 3) ; (1, 2) ; (1, 3) ; (1, 4) ; (1, 4) ; (1) ; (2, 3) ; (2, 4) ; (2, 4) ; (2) ; (3, 4) ; (3, 4) ; (3).

Placing of 5 in the above pattern is also unique and yields

(1, 2, 3, 4) ; (1, 2, 3, 5) ; (1, 2) ; (1, 3) ; (1, 4, 5) ; (1, 4) ; (1, 5) ; (2, 3) ; (2, 4, 5) ; (2, 4) ; (2, 5) ; (3, 4, 5) ; (3, 4) ; (3, 5).

Complete the fifth block with 6, and then 6 is to be placed in six two-treatment blocks (i.e. six blocks containing two treatments each). There are only two non-isomorphic ways of selecting these, viz.,

[α] (3, 5) ; (3, 4) ; (2, 5) ; (2, 3) ; (1, 4) ; (1, 2).
[β] (3, 5) ; (3, 4) ; (2, 5) ; (2, 4) ; (1, 3) ; (1, 2).

Thus we get two types of patterns :

[α] (1, 4, 5, 6) ; (1, 2, 3, 5) ; (1, 2, 3, 4) ; (3, 5, 6) ; (3, 4, 6) ; (3, 4, 5) ; (2, 5, 6) ; (2, 4, 5) ; (2, 3, 6) ; (1, 4, 6) ; (1, 2, 6) ; (2, 4) ; (1, 5) ; (1, 3).
[β] (1, 4, 5, 6) ; (1, 2, 3, 5) ; (1, 2, 3, 4) ; (3, 5, 6) ; (3, 4, 6) ; (3, 4, 5) ; (2, 5, 6) ; (2, 4, 6) ; (2, 4, 5) ; (1, 3, 6) ; (1, 2, 6) ; (2, 3) ; (1, 5) ; (1, 4).

Completing the third block with 7, we find that 7, which is to be placed in three three-treatment blocks and three two-treatment blocks, can be placed in the following blocks in [α] and [β] types :

Case of [α]
[α_1] (3, 5, 6) ; (3, 4, 6) ; (2, 4, 6) ; (1, 2, 6) ; (2, 4) ; (1, 5) ; (1, 3).
[α_2] (3, 4, 6) ; (3, 4, 5) ; (2, 5, 6) ; (1, 2, 6) ; (2, 4) ; (1, 5) ; (1, 3).

Case of [β]
[β_1] (3, 5, 6) ; (3, 4, 6) ; (2, 4, 5) ; (1, 2, 6) ; (2, 3) ; (1, 5) ; (1, 4).
[β_2] (3, 5, 6) ; (3, 4, 5) ; (2, 4, 6) ; (1, 2, 6) ; (2, 3) ; (1, 5) ; (1, 4).

But out of these, placing of 7 in [β_1] is isomorphic with that of 6 in [α]. Hence we get three solutions, starting from an initial block which follows law (a) :

[α_1] (3, 5, 6, 7) ; (3, 4, 6, 7) ; (3, 4, 5, 8) ; (2, 5, 6, 8) ; (2, 4, 7, 8) ; (2, 4, 5, 7) ;
(2, 3, 6, 8) ; (1, 5, 7, 8) ; (1, 4, 6, 8) ; (1, 4, 5, 6) ; (1, 3, 7, 8) ; (1, 2, 6, 7) ;
(1, 2, 3, 5) ; (1, 2, 3, 4).

[α_2] (3, 5, 6, 8) ; (3, 4, 6, 7) ; (3, 4, 5, 7) ; (2, 5, 6, 7) ; (2, 4, 7, 8) ; (2, 4, 5, 8) ;
(2, 3, 6, 8) ; (1, 5, 7, 8) ; (1, 4, 6, 8) ; (1, 4, 5, 6) ; (1, 3, 7, 8) ; (1, 2, 6, 7) ;
(1, 2, 3, 5) ; (1, 2, 3, 4).

[β_1] (3, 5, 6, 7) ; (3, 4, 6, 7) ; (3, 4, 5, 8) ; (2, 5, 6, 8) ; (2, 4, 6, 8) ; (2, 4, 5, 7) ;
(2, 3, 7, 8) ; (1, 5, 7, 8) ; (1, 4, 7, 8) ; (1, 4, 5, 6) ; (1, 3, 6, 8) ; (1, 2, 6, 7) ;
(1, 2, 3, 5) ; (1, 2, 3, 4).

If we look into the solutions minutely, the following characteristics of the above three designs will be revealed :

[α_1] Twelve blocks follow law (a) and two follow law (b)
[α_2] All the fourteen blocks follow law (a)
[β_1] Eight blocks follow law (a) and six follow law (b).

This proves the non-isomorphism of the three solutions obtained and leads us to seek a solution where all the blocks follow law (b). There is really a single solution like this :

[γ] (1, 2, 3, 4) ; (1, 2, 5, 6) ; (1, 2, 7, 8) ; (1, 3, 5, 7) ; (1, 3, 6, 8) ; (1, 4, 5, 8) ;
(1, 4, 6, 7) ; (2, 3, 5, 8) ; (2, 3, 6, 7) ; (2, 4, 5, 7) ; (2, 4, 6, 8) ; (3, 4, 5, 6) ;
(3, 4, 7, 8) ; (5, 6, 7, 8).

[γ] is obviously non-isomorphic with the other three solutions and hence in all we get four solutions [α_1], [α_2], [β_1], [γ] for the configuration : $v=8$, $b=14$, $r=7$, $k=4$, $\lambda=3$.

The blocks of the derived configuration : $v=7$, $b=14$, $r=6$, $k=3$, $\lambda=2$, obey the following laws :

(a′) $f_0=1,\ \ f_1=9,\ \ f_2=3,\ \ f_3=0$

(b′) $f_0=0,\ \ f_1=12,\ \ f_2=0,\ \ f_3=1.$

Proceeding in the above lines, we get the four solutions :

[α'_1] (9, 10, 11) ; (9, 10, 12) ; (9, 11, 13) ; (10, 11, 14) ; (9, 12, 15) ; (9, 13, 14) ;
(9, 14, 15) ; (10, 12, 14) ; (10, 13, 15) ; (10, 13, 15) ; (11, 12, 13) ; (11, 12, 15) ;
(11, 14, 15) ; (12, 13, 14).

[α'_2] (9, 10, 11) ; (9, 10, 12) ; (9, 11, 13) ; (10, 11, 14) ; (9, 12, 14) ; (9, 13, 15) ;
(9, 14, 15) ; (10, 12, 15) ; (10, 13, 14) ; (10, 13, 15) ; (11, 12, 13) ; (11, 12, 15) ;
(11, 14, 15) ; (12, 13, 14).

[β'_1] (9, 10, 11) ; (9, 10, 12) ; (9, 11, 13) ; (10, 11, 14) ; (9, 12, 13) ; (9, 14, 15) ;
(9, 14, 15) ; (10, 12, 14) ; (10, 13, 15) ; (10, 13, 1t) ; (11, 12, 15) ; (11, 12, 15) ;
(11, 13, 14) ; (12, 13, 14).

[γ'] (9, 10, 11) ; (9, 10, 11) ; (9, 12, 13) ; (9, 12, 13) ; (9, 14, 15) ; (9, 14, 15) ;
(10, 12, 14) ; (10, 12, 14) ; (10, 13, 15) ; (10, 13, 15) ; (11, 12, 15) ; (11, 12, 15) ;
(11, 13, 14) ; (11, 13, 14) ;

These four solutions are characterised by the following properties :

[α_1'] Twelve blocks obey law (a′) and two (b′)
[α'_2] All blocks obey law (a′)
[β'_1] Eight blocks obey law (a′) and six law (b′)
[γ'] All blocks obey law (b′).

3. ENUMERATION OF THE SYMMETRICAL CONFIGURATION

It is obvious that all symmetrical configurations can be obtained from the suitable adjunction of blocks of derived configuration to those of the residual configuration with the addition of a new block containing the treatments involved in the derived configuration. But for adjunction, the following necessary condition must be satisfied :—if 't' blocks of the residual configuration follow law (a) and the rest obey law (b), then in the derived configuration to be adjoined to it, 't' blocks must obey the complementary law (a′) and the rest, the complementary law (b′). Hence we must try the adjunction of (i)[α_1] and [α'_1], (ii) [α_2] and [α'_2], (iii) [β_1] and [β'_1] and (iv) [γ] and [γ'].

Let us consider the first case. There are two different ways of adjunction as shown below, all other ways of adjunction being obtained from these two by simply renaming the blocks and treatments of the adjoined derived configuration. The method followed is like this :

$[\alpha_1]$	$[\alpha'_1]_1$	$[\alpha'_1]_2$
(1, 2, 3, 4)	(9, 10, 11)	(9, 10, 11)
(1, 2, 3, 5)	(12, 13, 14)	(12, 13, 14)
(1, 2, 6, 7)	(10, 13, 15)	(10, 13, 15)
(1, 3, 7, 8)	(9, 12, 15)	(11, 14, 15)
(1, 4, 5, 6)	(11, 12, 15)	(11, 12, 15)
(1, 4, 6, 8)	(9, 13, 14)	(9, 13, 14)
(1, 5, 7, 8)	(10, 11, 14)	(9, 10, 12)
(2, 3, 6, 8)	(11, 14, 15)	(9, 12, 15)
(2, 4, 5, 7)	(9, 14, 15)	(9, 14, 15)
(2, 4, 7, 8)	(11, 12, 13)	(11, 12, 13)
(2, 5, 6, 8)	(9, 10, 12)	(10, 11, 14)
(3, 4, 5, 8)	(10, 13, 15)	(10, 13, 15)
(3, 4, 6, 7)	(10, 12, 14)	(10, 12, 14)
(3, 5, 6, 7)	(9, 11, 13)	(9, 11, 13)

{9, 10, 11, 12, 13, 14, 15}

To the first block of $(\alpha_1]$ which follows law (a) any block of $[\alpha_1']$ which follows law (a) is associated. For the second block of $[\alpha_1]$ we must choose a block of $[\alpha'_1]$ such that the two blocks, after adjunction, have just three treatments in common. This procedure is repreated ; if at any stage, more than one blocks of $[\alpha'_1]$ can be chosen, it is seen whether a renaming of the treatments of $[\alpha'_1]$ makes the differently chosen blocks identical, keeping the already adjoined blocks the same and the whole $[\alpha'_1]$ the same.

In the same way it ihas been found that $[\alpha'_2]$ can be adjoined to $[\alpha_2]$ in one manner only :

$[\alpha_2]$	$[\alpha'_2]$
(1, 2, 3, 4)	(9, 10, 11)
(1, 2, 3, 5)	(12, 13, 14)
(1, 2, 6, 7)	(9, 13, 15)
(1, 3, 7, 8)	(11, 14, 15)
(1, 4, 5, 6)	(11, 12, 15)
(1, 4, 6, 8)	(10, 13, 14)
(1, 5, 7, 8)	(9, 10, 12)
(2, 3, 6, 8)	(10, 12, 15)
(2, 4, 5, 8)	(9, 14, 15)
(2, 4, 7, 8)	(11, 12, 13)
(2, 5, 6, 7)	(10, 11, 14)
(3, 4, 5, 7)	(10, 13, 15)
(3, 4, 6, 7)	(9, 12, 14)
(3, 5, 6, 8)	(9, 11, 13)

{9, 10, 11, 12, 13, 14, 15}

$[\beta_1]$ and $[\beta'_1]$ can be adjoined in six different ways but except the first which is given below, the other five adjunctions lead to configurations isomorphic with $[\alpha_2\alpha'_2]$ and $[\alpha_1\alpha'_1]_2$ already obtained, as is easily seen that if the first completed block is taken as the initial configuration, the derived configuration is $[\alpha'_1]$ or $[\alpha'_2]$.

$[\beta_1]$	$[\beta'_1]$
(1, 2, 3, 4)	(9, 10, 11)
(1, 2, 3, 5)	(12, 13, 14)
(1, 2, 6, 7)	(10, 13, 15)
(1, 3, 6, 8)	(11, 12, 15)
(1, 4, 5, 6)	(9, 14, 15)
(1, 4, 7, 8)	(9, 12, 13)
(1, 5, 7, 8)	(10, 11, 14)
(2, 3, 7, 8)	(9, 14, 15)
(2, 4, 5, 7)	(11, 12, 15)
(2, 4, 6, 8)	(11, 13, 14)
(2, 5, 6, 8)	(9, 10, 12)
(3, 4, 5, 8)	(10, 13, 15)
(3, 4, 6, 7)	(10, 12, 14)
(3, 5, 6, 7)	(9, 11, 13)

{9, 10, 11, 12, 13, 14, 15}

$[\gamma]$ and $[\gamma']$ admit of twelve different adjunctions but except the one given below, all leads to configurations isomorphic with $[\alpha_1\alpha'_1]_1$, $[\alpha_1\alpha'_1]_2$, $[\beta_1\beta'_1]$.

$[\gamma]$	$[\gamma']$
(1, 2, 3, 4)	(9, 10, 11)
(1, 2, 5, 6)	(9 , 12, 13)
(1, 2, 7, 8)	(9, 14, 15)
(1, 3, 5, 7)	(10, 12, 14)
(1, 3, 6, 8)	(10, 13, 15)
(1, 4, 5, 8)	(11, 12, 15)
(1, 4, 6, 7)	(11, 13, 14)
(2, 3, 5, 8)	(11, 13, 14)
(2, 3, 6, 7)	(11, 12, 15)
(2, 4, 5, 7)	(10, 13, 15)
(2, 4, 6, 8)	(10, 12, 14)
(3, 4, 5, 6)	(9, 14, 15)
(3, 4, 7, 8)	(9, 12, 13)
(5, 6, 7, 8)	(9, 10, 11)

{9, 10, 11, 12, 13, 14, 15}

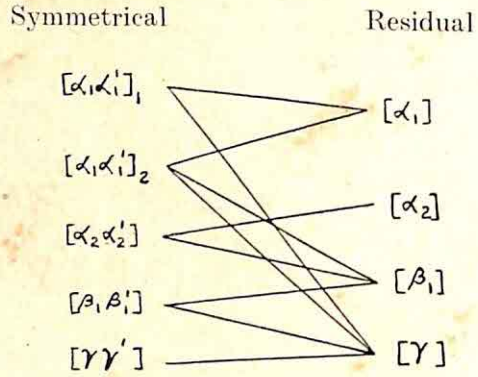
All the five symmetrical configurations, viz., $[\alpha_1\alpha'_1]_1$, $[\alpha_1\alpha'_1]_2$, $[\alpha_2\alpha'_2]$, $[\beta_1\beta'_1]$, $[\gamma\gamma']$ are non-isomorphic as can be seen by considering all possible derived configurations of each.

$(\alpha_1\alpha'_1)_1$ gives fourteen derived configurations $[\alpha'_1]$ and one $[\gamma']$.
$(\alpha_1\alpha'_1)_2$ gives eight $[\alpha_1']$, six $[\beta'_1]$, and one $[\gamma']$
$(\alpha_8\alpha'_1)$ gives eight $[\alpha'_2]$ and seven $[\beta'_1]$
$(\beta_8\beta'_1)$ gives twelve $[\beta'_1]$ and three $[\gamma']$
$(\gamma\gamma')$ gives fifteen $[\gamma']$

4. CONCLUSIONS

The symmetrical configuration : $v=b=15$, $r=k=7$, $\lambda=3$ has five non-isomorphic solutions given in the paper as $[\alpha_1\alpha'_1]_1$, $[\alpha_1\alpha'_1]_2$, $[\alpha_2\alpha'_2]$, $[\beta_1\beta'_1]$ and $[\gamma\gamma']$. The solutions for the same given in Fisher and Yates' table (1938) and Bose' paper (1939) are isomorphic with one another and with $[\gamma\gamma']$ of this paper. The residual configuration as well as the derived configuration has each four non-siomorphic solutions given in the paper respectively as $[\alpha_1]$, $[\alpha_2]$, $[\beta_1]$, $[\gamma]$ and $[\alpha'_1]$, $[\alpha'_2]$, $[\beta'_1]$, $[\gamma']$. The solutions given for the residual configuration by Fisher and Yates (1938) and Bose (1939) are isomorphic with the solution $[\gamma]$ of this paper.

The most interesting point is the relation between the symmetrical configuration studied and its residuals or derivatives. The correspondence can be shown diagrammatically like this :



The bond in the above diagram indicates that we can pass from one to another. It reveals that from a single symmetrical design, we can obtain more than one non-isomorphic residual or derived configurations, and also that from two non-isomorphic symmetrical configurations we can obtain the identical residual or derivative if the initial blocks are chosen properly i.e. there is no one-to-one correspondence, in general, between the symmetrical configurations and their residuals or denrivatives.

REFERENCES

1. AYYANGER, A. A. KRISHNASWAMY (1943) A general theory of tactical configurations. *Jr. Mysore University*, 3, 103-113.

2. BOSE, R. C. (1939) On the construction of balanced incomplete block, design. *Ann. Eug.*, 9, 358-399.

3. COX, GERTRUDE, M. (1940) Enumeration and construction of balanced incomplete configurations. *Ann. Math. Stat.*, 9, 72-85.

4. FISHER, R. A. (1940) An Examination of the different possible solution of a problem in incomplete block. *Ann. Eug.*, 10, 52 75.

5. FISHER & YATES (1938) Stat. Tables for Biolocgical, Agricultural, and Medical Research.

6. HUSAIN, Q. M. (1944) On the Totality of solutions of symmetrical incomplete block designs. *Proc. Ind. Sci. Congress.*

7. NANDI, H. K. (1944) Note on the enumeration of all possible non-isomorphic solutions of balanced incomplete block designs. *Proc. Ind. Sci. Congress.*

Paper received 29 August, 1945

IMPOSSIBILITY OF THE SYMMETRICAL INCOMPLETE BLOCK DESIGN WITH $\lambda=2$, $k=7$

By Q. M. HUSAIN
Dacca University

In Fisher and Yates' statistical tables (1938, 1943) the authors have incorporated a number of balanced incomplete block designs for values of v, b, r, k and λ such that $k \leqslant 10$, where v, b, r, k and λ stand for the respective numbers of varieties, blocks, replications of each variety, varieties in each block and replication of each pair of varieties. Solutions to a few of the unsolved designs have been given by R. C. Bose, but there are still others for which no solutions have been found. These designs must satisfy the two necessary but not sufficient conditions $bk=vr$ and $r(k-1)=\lambda(v-1)$. Several sets of parameters satisfying the above conditions can be found but the existence or non-existence of corresponding solutions have not been proved in all cases. The set $v=b=22$, $r=k=7$, $\lambda=2$ has not yielded a solution so far and it is not known whether a solution at all exists. It is the object of the present paper to establish the non-existence of the design by a process of systematic exhaustion. In an earlier paper (Husain, 1945) the same method has been used to prove that the symmetrical design $\lambda=2$, $k=5$ admits of only one solution, and that the symmetrical design $\lambda=2$, $k=6$ admits of three non-isomorphic solutions.

The method followed is to build up the pattern gradually in a standard form. This process is bound to give a solution if it really exists. Further the tedious process of checking up a solution to see if it actually satisfies the required conditions is avoided, as the design, if it can be completed in the manner suggested must automatically be a solution. The design is built up as follows.

(1) In the first row, representing the first block, put the seven leading numbers 1, 2, 3, 4, 5, 6, 7 to represent the varieties in this block.

(2) As each of the other blocks must contain one of the pairs of these seven varieties, write down the twenty-one number pairs (1, 2), (1, 3), ..., (5, 6), (5, 7) and (6, 7) in the standard order in the remaining twenty one blocks.

(3) Complete the second row by putting down the next five numbers 8, 9, 10, 11 and 12 in order. The first two rows and columns are thus written down in a standard form.

(4) The five varieties 8, 9, 10, 11 and 12 are now to be placed six times each in the remaining twenty blocks. The variety 8 must occur in conjunction with each of the seven varieties already placed. The placing of variety 8 can be conveniently represented by means of cycles which may be called λ-chains. Thus 8 : (1 2 3 4 5 6 7) means that variety 8 is placed in the blocks containing the variety-pairs, (1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 7) and (7, 1). It is clear that for variety 8, only three possible λ-chains exist; 8 : (1 2 3 4 5 6 7), 8 : (1 2 3) (4 5 6 7) and 8 : (1 2 3 4) (5 6 7) where in the case of the two broken cycles the λ-chaining is performed separately for each of the smaller cycles. Thus 8 : (1 2 3) (4 5 6 7) means that, variety 8 is placed with the variety-pairs (1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 7) and (7, 4). We may denote these three standard placings of variety 8 as types A, B and C.

(5) Variety 9 has already been placed in conjunction with the three varieties 1, 2 and 8. The remaining blocks in which variety 9 can be placed contain either two or three varieties. As variety 9 must occur twice in conjunction with each of the varieties from 1 to 8 that have already been placed, we can determine the number (x) of two-variety blocks and the number (y) of three-variety blocks in which variety 9 should be placed as the solutions of the simultaneous equations $x+y=6$, $2x+3y=8\times 2-3=13$ which gives $x=5$ and $y=1$. The possible placings of variety 9 are obtained by trial.

(6) To place variety 10, we find that the blocks in which it can be placed contain either two or three varieties.

The appropriate simultaneous equations are $x+y=6$; $2x+3y=14$; giving $x=4$, $y=2$.

The blocks eligible for containing this variety are those that have not already two varieties in common with the second block where it has already been placed.

(7) Varieties 11 and 12 are to be disposed of in the same way if possible. At this stage it will be found that the remaining twenty blocks fall into two groups of ten each, containing three or four varieties respectively. This is evident from the fact that the values of (y) for the varieties 8, 9, 10, 11 and 12 are 0, 1, 2, 3 and 4 respectively giving in all ten blocks containing four varieties each.

(8) The third block is now completed by the four varieties 13, 14, 15, and 16 and these have to be placed six times each in the remaining nineteen blocks. On trial it will be found that it is impossible to place variety 13 in accordance with the requirements. The actual working of the method is shown below.

Type 4 : The set of leading numbers with the placing of 8 in the λ-chain 8 : (1 2 3 4 5 6 7) is shown in Chart 1.

CHART 1

1	2	3	4	5	6	7
1	2	8	9	10	11	12
1	3					
1	4					
1	5					
1	6					
1	7	8	×			
2	3	8	×			
2	4					
2	5					
2	6					
2	7					
3	4	8				
3	5					
3	6					
3	7					
4	5	8				
4	6					
4	7					
5	6	8				
5	7					
6	7	8				

CHART 2

(0)	(a)	35·46	36·46	37·46	35·47	36·47	37·47
12 34 57	13 24	×					
	14 24	×					
	15 24	×					
	16 24	×					
	13 25	×					
	14 25	×					
	15 25	×					
	16 25	×					
	13 26	×					
	14 26	×					
	15 26	×					
	16 26	×					
	13 27	×					
	14 27	×					
	15 27	×					
	16 27	×					

Evidently any other placing of variety 8 in a λ-chain of seven elements is merely a permutation of the leading numbers. Blocks in which variety 9 cannot be placed are indicated by a cross. These are blocks which already contain two varieties in common in the second block. It is seen that variety 9 can be placed in any one of the four blocks containing the leading pairs (3, 4), (4, 5), (5, 6) and (6, 7). Now it can be proved that the first and fourth as also the second and third ways of placing are isomorphous. Thus the substitution $\begin{pmatrix} 1\,2\,3\,4\,5\,6\,7 \\ 6\,7\,1\,2\,3\,4\,5 \end{pmatrix}$ shows that placing variety 9 in the block containing the variety pair (6, 7) is the same as placing it in the block with variety-pair (3, 4). Similarly the substitution $\begin{pmatrix} 1,\,2\,3\,4\,5\,6\,7 \\ 5\,6\,7\,1\,2\,3\,4 \end{pmatrix}$ shows the equivalence of placing variety 9 in the two blocks containing variety-pairs (5 6) and, (4, 5). We need therefore, examine the placings of variety 9 in the blocks containing (3, 4) and (4, 5) only.

On trial, it is clear that there are exactly thirteen ways of placing variety 9 which are shown below

$9 : (1\,2\,5\,7\,3\,4\,6)_1 : (1\,2\,5\,7\,4\,3\,6)_2 ; (1\,2\,6\,4\,3\,7\,5)_3 ; (1\,2\,6\,3\,4\,7\,5)_4 ; (1\,2\,6)\,(3\,4\,7\,5)_5 ;$
$(1\,2\,7\,5)\,(3\,4\,6)_6 ; (1\,2\,7\,5\,3\,4\,6)_7 ; (1\,2\,6\,4\,5\,7\,3)_8 ; (1\,2\,6\,3)\,(4\,5\,7)_9 : (1\,2\,6)\,(3\,5\,4\,7)_{10} ;$
$(1\,2\,7\,5\,4\,6\,3)_{11} : (1\,2\,7\,3\,5\,4\,6)_{12} ; (1\,2\,7\,4\,5\,3\,6)_{13}.$

The most convenient trial chart is Chart 2 (given above) based on the considerations that having placed variety 9 with varieties 1, 2, 8 and 3, 4, 8, we cannot use it with variety 8 any more, and that it must be used once with each of the varieties 1, 2, 3 and 4 and that the 5th placing must be with the pair (5, 7)not containing these varieties.

Chart 2 includes all possible conditions in which 1, 2, 3 and 4 can be taken once. Looking down the first column we notice that 12. 34. 57 in conjunction with 35. 46 includes the numbers 3, 4, 5 twice; so it cannot be used with any row of column (a) that contains these numbers. The only possible combination shown by an asterik is 12. 34. 57. 16. 27. 35. 46 which is 9 : (1 2 7 5 3 4 6) being the 7th placing in our list. In a similar manner, all the 13 possible placings were obtained. Now for each way of placing 9, the possible placings of 10 were obtained. The placings upto 9, and the appropriate chart for 9_1 10 are shown below in Chart 3.

CHART 3

9_1

1 2 3 4 5 6 7
1 2 8 9 10 11 12
1 3
1 4
1 5
1 6 9 ×
1 7 8 ×
2 3 8 ×
2 4
2 5 9 ×
2 6
2 7
3 4 8 9 ×
3 5
3 6
5 7 9
4 5 8
4 6 9
4 7
5 6 8
5 7 9
6 7 8

9,10 :

(0)	(a)	35·36	35·47	36·47	remarks
	45·37	×			One 8 and one 9 are to be placed. This explains column (a). Cut of 4 double numbers, two must contain 1 and 2 and two must come from doublets not containing 1 or 2. This explains the tops of the other columns.
	56·37	×			
	67·37	×			
	45·46	(12) 77			
	56·46	×			
	56·46	×			
	45·57	×			
	56·57	×			
	67·57	(12) 44			

The obvious failures are indicated by crosses. In the others, the missing numbers are written down—i.e. the numbers which would give on the whole each of the numbers from 1 to 9 twice over. In the present case, (17, 27) is not possible but (14, 24) is possible giving the successful placing as 12. 67. 57. 35. 36. 14. 24 or 10 : (1 2 4) (3 5 7 6)

I have found this to be quickest way of trying out the combinations of 10. The full set of combinations of 9_1 10 is :—

9_1 10 : (1 2 4) (3 5 7 6) ; (1 2 4 7 5 6 3) ; (1 2 7 4 6 5 3) ; (1 2 6 5 3 7 4) ; (1 2 4 7 3 6 5) ;
(1 2 6 3) (4 5 7) ; (1 2 6 3 7 4 5).

The pattern 9_1 10_1 is shown below in Chart 4, together with the trial chart for placing 11. Note that in this case, $x=3$, $y=3$.

CHART 4.

$9_1 10_1$

1 2 3 4 5 6 7
1 2 8 9 10 11 12
1 3
1 4 10 ×
1 5
1 6 9 ×
1 7 8 ×
2 3 8 ×
2 4 10 ×
2 5 9 ×
2 6
2 7
3 4 8 9 ×
3 5 10
3 6 10
3 7 9
4 5 8
4 6 9
4 7
5 6 8
5 7 9 10 ×
6 7 8 10 ×

(o)	(a)	13·26	13·27	15·26	15·27
47·12 (8 9 10)	35 37 45	×	×	×	×
	36 37 45	×	×	×	×
	35 46 45	×	×	×	×
	36 46 45	×	×	×	×
	35 37 56	×	×	×	×
	36 37 56	×	×	×	×
	35 46 56	×	×	×	×
	36 46 56	×	×	×	×

In this chart column (a) is written to embrace all combinations in which one 8, one 9 and one 10 can be taken. The other parts need no explanation. There are only two successful placings for 11. If 11 is placed in one of the ways, 12 must be placed in the other way. The placings are 9_1 10_1 11 : (47. 12. 36. 37. 45. 15. 26) ; (47, 12, 35, 46, 56, 13. 27) or 9, 10, 11 : (1 2 6 3 7 4 5) ; (1 2 7 4 6 5 3). When the second block is completed and all the varieties in this block are suitably placed, the pattern is as shown in Chart 5 and 6 :—

CHART 5.

9_1 10_1 11_1 12_1						
1	2	3	4	5	6	7
1	2	8	9	10	11	12
1	3	12	13	14	15	16
1	4	10				
1	5	11				
1	6	9				
1	7	8				
2	3	8				
2	4	10				
2	5	9				
2	6	11				
2	7	12				
3	4	8	9			
3	5	10	12			
3	6	10	11			
3	7	9	11			
4	5	8	11		p	
4	6	9	12			
4	7	11	12			
5	6	8	12			
5	7	9	10		q	
6	7	8	10			

CHART 6.

Trial Chart for 9_1 10_1 11_1 13 :

(o)	(a)		14 10.	15 11.	16 9.	17 8	4-ers
1312	2 7 12	2 3 8	×	×	×	×	three
	4 6 9 12	2 3 8	×	×	×	×	two
	4 7 11 12	2 3 8	×	×	(2610)?	×	two
	5 6 8 12	2 8 3	×	×	×	×	two
	2 7 12	3 4 8 9	×	×	×	×	two
	4 6 9 12	3 4 8 9	×	(22 5 7 10 11)	×	×	one
	4 7 11 12	3 4 8 9	×	(22 5 6 9 10)	×	×	one
	5 6 8 12	3 4 8 9	×	×	×	×	one
	2 7 12	3 6 10 11	×	×	(248)?	×	two
	4 6 9 12	3 6 10 11	×	×	×	(2257910 2445811)	one
	4 7 11 12	3 6 10 11	×	×	×	×	one
	5 6 8 12	3 6 10 11	×	×	×	×	one
	2 7 12	3 7 9 11	×	×	×	×	two
	4 6 9 12	3 7 9 11	×	(2245810)	×	×	one
	4 7 11 12	3 7 9 11	×	×	×	×	one
	5 6 8 12	3 7 9 11	(2267910 2246811 2245911)	×	×	×	one

Now, we have to place 13, 14, 15 and 16 in the third row and the remaining rows that will take them.

We notice, $x+y=6$, $3x+4y=24-3=21$, which gives $x=3$, $y=3$. On trial, we find it impossible to place 13.

Combinations that obviously fail, are underlined. Some of them can accomodate all the chart and need the numbers shown in brackets. Every time it is seen that no such combination is possible.

Thus in this test case it is impossible to place 13. This has been found to be the case in all the tests that exhaust the possibilities. To facilitate checking by any future investigator the results of the trials of all the possibilities are listed below for reference :—

Type A=8 : (1 2 3 4 5 6 7)

9_1 : (1 2 5 7 3 4 6) ; (1 2 5 7 4 3 6) ; (1 2 6 4 3 7 5) ; (1 2 6 3 4 7 5) ; (1 2 6) (3 4 7 5) ; (1 2 7 5)(3 4 6) ;
(1 2 7 5 3 4 6) ; (1 2 6 4 5 7 3) ; (1 2 6 3) (4 5 7) ; (1 2 6) (3 5 4 7) ; (1 2 7 5 4 6 3) ;
(1 2 7 3 5 4 6) ; (1 2 7 4 5 3 6) ;
$9_1 10$: (1 2 4) (3 5 7 6) ; (1 2 4 7 5 6 3) ; (1 2 7 4 6 5 3) ; (1 2 6 5 3 7 4) : (1 2 4 7 3 6 5) :
(1 2 6 3) (4 5 7) ; (1 2 6 3 7 4 5).
$9_1 10_1 11$: (1 2 6 3 7 4 5) ; (1 2 7 4 6 5 3)
$9_1 10_2$: fails to place 13
$9_1 10_3 11$: (1 2 4) (3 5 7 6) ; (1 2 6 3 7 4 5) fails to place 11
$9_1 10_4$: fails to place 13
$9_1 10_5$: fails to place 11
$9_1 10_6$: fails to place 11
$9_1 10_7$ 11 : (1 2 7 4 6 5 3) ; (1 2 4) (3 5 7 6) fails to place 11
fails to place 13.

The three combinations, successful up to 12, are really the same, being only different permutations of 10, 11 and 12.

$9_2 10$: (1 2 4 6 5 7 3) ; (1 2 4 6 7 5 3) ; (1 2 6 4 5 7 3) ; (1 2 6 5 3 7 4) ; (1 2 7 3 6 4 5)
$9_2 10_1$: fails to place 11
$9_2 10_2 11$: (1 2 7 3 6 4 5) ; (1 2 6 5 3 7 4) fails to place 13
$9_2 10_3$: fails to place 11

$9_2 10_4$ and $9_2 10_5$ are simply variants of $9_2 10_2$.

$9_3 10$: (1 2 4 7 5 6 3) ; (1 2 5 3) (4 6 7) ; (1 2 7 4 6 5 3) ; (1 2 4) (3 5 7 6) ; (1 2 5 6 3 7 4)

(1 2 4 7 3 5 6) ; (1 2 5 4 7 3 6).

$9_3 10_1$: fails to place 11

$9_3 10_2$: fails to place 11

$9_3 10_3 11$: (1 2 4) (3 5 7 6) ; (1 2 5 4 7 3 6) fails to place 13

$9_3 10_4$: is a variant of $9_3 10_2$.

$9_3 10_5$: fails to place 11

$9_3 10_6$: fails to place 11

$9_3 10_7$: is a variant of $9_3 10_3$

$9_4 10$: (1 2 4 6 5 7 3) ; (1 2 4 6 7 5 3) ; (1 2 5 6 4 7 3) ; (1 2 5 3) (4 6 7) ;

(1 2 7 4 6 5 3) ; (1 2 4 7 3 5 6) all these fail to place 11

$9_5 10$: (1 2 5 6 3 7 4) ; (1 2 4 7 3 6 5) ; (1 2 4 3 6 5 7) ; (1 2 7 5 4 6 3) ; (1 2 7 3 5 6 4) ; (1 2 5 6 4 7 3)

(1 2 4 6 5 7 3) ; (1 2 5 3 7 6 4) ; (1 2 4 6 7 3 5).

$9_5 10_1 11$: (1 2 7 5 4 6 3) ; (1 2 4 6 7 3 5) fails to place 13

$9_5 10_2 11$: (1 2 7 5 4 6 3) ; (1 2 5 3 7 6 4) fails to place 13.

These show that 10_1, 10_8, 10_9 must fail to place 13. The rest fail to place 11.

$9_6 10$: (1 2 5 3 7 6 4) ; (1 2 5 3 6 7 4) ; (1 2 5 6 3 7 4) ; (1 2 5 3) (4 6 7) ;

(1 2 5 6 4 7 3) ; (1 2 5 4 7 3 6) All these fail to place 11

$9_7 10$: (1 2 6 3 7 5 4) ; (1 2 4) (3 6 5 7) ; (1 2 5 6 4 7 3) ; (1 2 6 5 3 7 4) ; (1 2 4 7 5 6 3) ;

(1 2 6 3) (4 5 7) ; (1 2 5 3 6 7 4) ; (1 2 4 7 6 3 5)

$9_7 10_1 11$: (1 2 5 6 4 7 3) ; (1 2 4 7 6 3 5) fails to place 13

Therefore 10_2 and 10_8 also fail to place 13. All the rest fail to place 11.

$9_8 10$: (1 2 4) (3 5 7 6) ; (1 2 5 7 4 3 6) ; (1 2 5 6 3 7 4) ; (1 2 4 7 3 6 5) ;

(1 2 4 7 3 5 6) All these fail to place 11

$9_9 10$: (1 2 4 7 3 5 6) ; (1 2 7 5 3 4 6) ; (1 2 5 7 3 4 6) fails to place 11

$9_{10} 10$: (1 2 4 7 5 6 3) ; (1 2 4 6 5 7 3) ; (1 2 4 6 7 5 3) ; (1 2 4) (3 5 7 6) ; (1 2 4) (3 6 5 7)

All these fail to place 11.

$9_{11} 10$: (1 2 5 3 7 6 4) ; (1 2 4 6 7 3 5) ; (1 2 5 6 3 7 4) ; (1 2 4 7 3 6 5) ; (1 2 5 3 6 7 4) ;

(1 2 4 7 6 3 5) ; (1 2 6) (3 4 7 5).

$9_{11} 10_1 11$: (1 2 6) (3 4 7 5) ; (1 2 4 7 3 6 5) fails to place 13

$9_{11} 10_2 11$: (1 2 5 6 3 7 4) ; (1 2 6) (3 4 7 5) fails to place 13

therefore 10_4 also fails

therefore 10_7 also fails

therefore 10_3 also fails

10_5 and 10_6 fail to place 11.

$9_{12} 10$: (1 2 5 3 6 7 4) ; (1 2 5 6 3 7 4) ; (1 2 4 7 6 3 5) ; (1 2 4 7 3 6 5) ; (1 2 4) (3 5 7 6) ;

(1 2 4) (3 6 5 7) All these fail to place 11

$9_{13} 10$: (1 2 5 3 7 6 4) ; (1 2 4 6 7 3 5) ; (1 2 5 6 4 7 3) ; (1 2 4 6 7 5 3) ; (1 2 4) (3 6 5 7)

All these fail to place 11. Thus Type A fails.

As some of the placings of type A break up into smaller cycles of the types B and C, and the varieties 8 to 12 are interchangeable there is no essential difference between these types. Nevertheless, the distinction has been maintained as convenient. Also some of the patterns of these types are common (so far as varieties up to 12 are concerned). This affords an independent check for detecting whether a successful pattern has been left out.

Type B. 8 : (1 2 3 4) (5 6 7)

In this type, any standard pattern (completed upto variety 9) containing the combination 5 6 8 9 can be transformed to one containing either 6789 or 5789 as is obvious. So we need consider only the placings 3489 and 5689. The former does not give any pattern up to 9, so it is enough to consider only the placing 5689 for the variety 9.

9 : (1 2 4 7 3 6 5) ; (1 2 4 7 3 5 6) ; (1 2 4 7) (3 5 6) ; (1 2 4 6 5 3 7) ; (1 2 4 5 6 3 7) ;

(1 2 7 4 5 6 3) ; (1 2 7 3) (4 5 6).

(1 2 5 6 4 7 3) ; (1 2 6 5 4 7 3) ; (1 2 7 4 6 5 3) ; (1 2 7 5 4 6 3) ; (1 2 5 3) (4 6 7) ;

$9_1 10$: (1 2 6 7 4 5 3) ; (1 2 5 4 6 7 3) ; (1 2 6 4 5 7 3) ;

(1 2 6 4 7 5 3) All these fail to place 11.

$9_2 10$: This is really the same as 9_1 with the transposition (56) which leaves the earlier placings intact. So it fails to place 11.

$9_3 10$: (1 2 7 5 4 6 3) ; (1 2 7 6 4 5 3) fails to place 11 (with less than 3 ways of 10, it is impossible).

$9_4 10$: (1 2 6 7 4 5 3) ; (1 2 7 5 4 6 3) ; (1 2 5 7 4 6 3) fails to place 11.

This is obvious from the consideration that the blocks (1, 2) and (1, 3) cannot contain three varieties in common.

$9_5 10$: The same as 9_4 with transposition (56) which leaves earlier placing in tact.

$9_6 10$: (1 2 4 5) (3 6 7) : (1 2 4 5 7 3 6) : (1 2 4 7 6 3 5) : (1 2 4 7 5 3 6) : (1 2 4 6 3 5 7) :
(1 2 4 5 3 7 6) All these fail to place 11

$9_7 10$: This is a transposition of 9_6 by (56). It fails to place 11.

$9_8 10$: (1 2 4 6 3 7 5) : (1 2 4 5 3 7 6) : (1 2 4 5 3 6 7).

This must fail to place 11 ; the blocks (1, 2) and (2, 4) must not contain more than two varieties in common.

$9_9 10$: This is the same as 9_8 by the transposition (56). It fails to place 11.

$9_{10} 10$: (1 2 4 5 3 6 7) : (1 2 4 6 3 5 7). It fails to place 11.

(With less than three placings for 10, it is impossible to place 11 and 12).

Thus, type B fails

Type C : 8 (1 2 3) (4 5 6 7)

In this type, the placings of 9 in the blocks (4, 5), (5, 6), (6, 7) and (7, 4) are equivalent. So we need consider only the placing of 4 5 8 9.

9 : (1 2 4 5 7 3 6) ; (1 2 6 3 7 5 4) ; (1 2 6) (3 4 5 7) ; (1 2 6 3 4 5 7) ; (1 2 6 4 5 3 7) :
(1 2 7 3 6 4 5) ; (1 2 7 5 4 3 6) : (1 2 7 3 5 4 6) ; (1 2 7) (3 5 4 6).

$9_1 10$: (1 2 6 7 5 3 4) ; (1 2 5) (3 4 6 7) : (1 2 7 3 4 6 5) : (1 2 5 6 4 3 7) : (1 2 5 3 6 4 7) :
(1 2 7 4 6 3 5) ; (1 2 6 4 7 3 5) ; (1 2 7 3 5 6 4) : (1 2 5 3 7 6 4).

$9_1 10_1 11$: (1 2 7 3 4 6 5) : (1 2 7 4 6 3 5) : (1 2 5 6 4 3 7) : (1 2 5 3 6 4 7).

All these fail to place 13. So, 10_3, 10_4, 10_5, 10_6 must also fail. The others, viz. 10_2, 10_7, 10_8, 10_9 fail to place 11.

$9_2 10$: (1 2 4 3 5 7 6) ; (1 2 5) (3 4 6 7) : (1 2 7 3 4 6 5) ; (1 2 5 6 4 3 7) ; (1 2 4 6 7 3 5) :
(1 2 4 6 5 3 7) : (1 2 5 3 7 4 6) ; (1 2 7 4 6 3 5) ; (1 2 5 3 6 4 7).

$9_2 10_1 11$: (1 2 7 3 4 6 5) ; (1 2 5 6 4 3 7) ; (1 2 7 4 6 3 5) ; (1 2 5 3 6 4 7).

They all fail to place 13. So 10_3, 10_4, 10_8 and 10_9 also fail. Also 10_2, 10_5, 10_6 and 10_7 fail to place 11.

$9_3 10$: (1 2 4) (3 5 7 6) ; (1 2 5 3 7 6 4) ; (1 2 4 6 7 3 5) : (1 2 7 3 5 6 4) ; (1 2 4 6 5 3 7) :
(1 2 7 6 4 3 5) ; (1 2 5 3 4 6 7) : (1 2 7) (3 4 6 5) ; (1 2 5) (3 6 4 7).

$9_3 10_1 11$: (1 2 5) (3 6 4 7) ; (1 2 7) (3 4 6 5) fail to place 13.

So, 10_8 and 10_9 also must fail

The other combinations 10_2, 10_3, 10_4, 10_5, 10_6, 10_7 fail to place 11.

$9_4 10$: (1 2 7 4 6 3 5) ; (1 2 7 6 4 3 5) : (1 2 7 3 4 6 5) : (1 2 5) (3 4 6 7) ; (1 2 5) (3 6 4 7)
All these fail to place 11.

$9_5 10$: (1 2 4 3 7 5 6) ; (1 2 4 3 5 7 6) : (1 2 4 7 5 3 6) : (1 2 4) (3 5 7 6) ; (1 2 4) (3 6 5 7)
All these fail to place 11.

$9_6 10$: (1 2 5 7 4 3 6) ; (1 2 4 3 6 5 7) ; (1 2 5 7 6 3 4) : (1 2 6 5 7 3 4) ; (1 2 4 3 7 5 6) :
(1 2 6 3 5 7 4) ; (1 2 4 7 5 3 6) ; (1 2 4) (3 5 7 6) ; (1 2 5 3 4 6 7).

In this case, 10_1, 10_2, 10_3, 10_8 fail to place 11.

$9_6 10_4 11$: (1 2 4 7 5 3 6) ; (1 2 5 3 4 6 7) fails to place 13

$9_6 10_5 11$: (1 2 6 3 5 7 4) ; (1 2 5 3 4 6 7) fails to place 13.

The failure of 10_4 and 10_5 shows that 10_6, 10_7 and 10_9 also must fail

$9_7 10$: (1 2 5 3 6 4 7) ; (1 2 5 3 4 6 7) ; (1 2 5 6 4 3 7) ; (1 2 5) (3 4 6 7) ; (1 2 5) (3 6 4 7)
All these fail to place 11.

$9_8 10$: (1 2 6 5 7 3 4) ; (1 2 6 7 5 3 4) : (1 2 6 3 5 7 4) ; (1 2 4) (3 5 7 6) ; (1 2 4) (3 6 5 7)
All these fail to place 11

$9_9 10$: (1 2 4) (3 6 5 7) ; (1 2 5) (3 4 6 7) ; (1 2 6) (3 4 7 5) ; (1 2 6 7 5 3 4) ; (1 2 4 3 5 7 6) :
(1 2 5 7 4 3 6) ; (1 2 6 3 4 7 5) : (1 2 4 3 6 7 5) ; (1 2 5 7 6 3 4).

$9_9 10_1 11$: (1 2 5) (3 4 6 7) ; (1 2 6) (3 4 7 5) fails to place 13. So, 10_2 and 10_3 also fail.
10_4, 10_5, 10_6, 10_7, 10_8 and 10_9 fail to place 11.

So type C fails.

It is thus definitely established that the parameters $v=22$, $b=22$, $r=7$, $k=7$, $\lambda=2$ admit of no corresponding design.

REFERENCES

BOSE, R. C. (1939) On the construction of balanced incomplete block designs. *Annals of Eugenics*, 9, 353-399.

FISHER, R. A. and YATES, F. (1938, 1943) Statistical Tables for Biological, Agricultural and Medical Research Oliver and Boyd (First and Second editions.) London and Edinburgh.

HUSAIN, Q. M. (1941) On the totality of solutions of the incomplete block designs $\lambda=2$, $k=5$ or 6.

Paper received 15 August, 1945

MISCELLANEOUS NOTES

ON THE ORDER OF APPROXIMATION INVOLVED IN LAPLACE'S CENTRAL LIMIT THEOREM IN PROBABILITY

By MANINDRA NATH GHOSH

Statistical Laboratory, Calcutta

The central limiting theorem in probability is often put forward as the final justification for the frequent use of the normal distribution. As far as the present author is aware no thorough investigations have, however, been made about the exact order of approximation involved. Professor Cramer (1937) has shown that when absolute moments upto third order exist for the summands the approximation is of the order of $1/\sqrt{n}$. This is the generalisation of a previous result of Liapounoff. However, although the central limiting theorem was proved to hold when moments of the order of $2+\delta (\delta>0)$ exist by Liapounoff (1900) and also under the more general conditions of Lindburgh no approximation formula was established which might make the theorem suitable for statistical and numerical purposes. The object of this paper is to secure an approximation formula for the case where absolute moments of the order of $(2+\delta)$ $(\delta>0)$ are assumed to exist.

In order to arrive at the desired approximation we shall follow the lines of proof by Kolmogoroff of the central limiting theorem as given by Khintchine (1933). Let $F_1(x)$ be a distribution function with mean zero and variance unity and let $x'_1, x'_2, \dots x'_n$ be n independant variates satisfying the same distribution law $F_1(x)$. For simplicity of analysis we shall introduce the variables $x_1=x'_1/\sqrt{n}, \dots, x_n=x'_n/\sqrt{n}$ satisfying the same distribution law $F_1(\sqrt{n}x)=F(x)$ with variance $1/n$ and mean zero. Denote by $U_k(x)$ the distribution function of the random variable $x=x_1+\dots+x_n$.
Then evidently we have

$$U_{k+1}(x)=\int_{-\infty}^{+\infty} U_k(x-\xi)\ dF(\xi) \qquad \dots (1)$$

We shall now show that $U_n(x)$ tends to the cumulative normal distribution function

$$\phi(x)=1/\sqrt{2\pi}\int_{-\infty}^{x} \exp(-t^2/2)\ dt$$

as n becomes large. The first part of the proof is a modification of the proof of Khintchine (1933) to suit our present limited purpose.

Let

$$\phi(x/\sqrt{z})=1/\sqrt{2\pi z}\int_{-\infty}^{x} \exp(-t^2/2z)dt$$

It is easy to see that $\phi(x/\sqrt{z})$ satisfies the differential equation $\partial\phi/\partial z=1/z.\ \partial^2\phi/\partial x^2$. Let us introduce the function $V(x,z)$ given by $V(x, z)=\phi(x/\sqrt{z})+\varepsilon z$ so that

$$\frac{\partial V}{\partial z}=\frac{1}{2}\frac{\partial^2 V}{\partial x^2}+\varepsilon \qquad \dots (2)$$

We shall now prove that for a suitably chosen $\varepsilon(\delta)$ in the whole halfplane $z>\delta$.

$$V(x, z+1/n)>\int_{-\infty}^{+\infty} V(x-\xi,z)\ dF(\xi)$$

Proof :

$$V(x-\xi, z) = V(x, z) - \xi\frac{\partial V}{\partial x}+\tfrac{1}{2}\xi^2\frac{\partial^2 V}{\partial x^2}+\tfrac{1}{2}\xi^2\rho(x,\xi,z) \qquad \dots (3)$$

where

$$\rho(x, \xi, z) = \left\{\left[\frac{\partial^2 V}{\partial x^2}\right]_{x-\theta\xi,z} - \left[\frac{\partial^2 V}{\partial x^2}\right]_{x, z}\right\}_{0<\theta<1}$$

$$\int_{-\infty}^{+\infty} V(x-\xi,z)\ dF(\xi)=V(x,z)+\frac{1}{2n}\frac{\partial^2 V}{\partial x^2}+\tfrac{1}{2}\int_{\infty}^{\infty}\xi^2\rho(x,\xi,z)\ dF(\xi)$$

as

$$\int_{-\infty}^{+\infty} dF(\xi)=1,\ \int_{-\infty}^{+\infty}\xi\ dF(\xi)=0 \text{ and } \int_{-\infty}^{+\infty}\xi^2\ dF(\xi) = 1/n.$$

From (2) it follows that if $\left| \int_{-\infty}^{+\infty} \xi^2 \, \rho(x,\xi,z) \, dF(\xi) \right| < \frac{\varepsilon}{n}$ .. (4)

then $$\int_{-\infty}^{+\infty} V(x-\xi,z) \, dF(\xi) < V(x,z) + \frac{1}{n} \frac{\partial V}{\partial z} - \frac{\varepsilon}{2n} \qquad .. \quad (5)$$

Again $$V\left(x,z+\frac{1}{n} \right) = V(x,z) + \frac{1}{n} \frac{\partial V}{\partial z} + \frac{1}{2n} \left[\frac{\partial^2 V}{\partial z^2} \right]_{x,z+\theta'/n, 0<\theta'<1} \qquad .. \quad (6)$$

Denoting by $M_1(\delta)$ the maximum value of $\frac{\partial^2 V}{\partial z^2}$ in the region $z>\delta$ when $M_1(\delta)/n<\varepsilon$ we have

$$V\left(x,z+\frac{1}{n} \right) > V(x,z) + \frac{1}{n} \frac{\partial V}{\partial z} - \frac{\varepsilon}{2n} \qquad .. \quad (7)$$

Hence from (5) and (7) we have

$$V\left(x,z+\frac{1}{n} \right) > \int_{-\infty}^{+\infty} V(x-\xi,z) \, dF(\xi) \qquad .. \quad (8)$$

provided ε is chosen to satisfy (4) and $M_1(\delta)/n<\varepsilon$

From (8) we have $\phi(x/\sqrt{(k+1)/n}) + \varepsilon(k+1)/n > \int_{-\infty}^{+\infty} \phi((x-\xi)/\sqrt{k/n}) \, dF(\xi) + k\varepsilon/n$ .. (9)

Now from (1) we have $U_{k+1}(x-2\alpha) = \int_{-\infty}^{+\infty} U_k(x-\zeta-2\alpha) \, dF(\xi)$ .. (10)

Combining these two results we have

$$U_{k+1}(x-2\alpha) - \phi(x/\sqrt{(k+1)/n}) < \frac{\varepsilon}{n} + \int_{-\infty}^{+\infty} [U_k(x-\xi-2\alpha) - \phi(x/\sqrt{k/n})] \, dF(\xi) \qquad .. \quad (11)$$

Denote by P_k the least upper bound of $U_k(x) - \phi\{(x+2\alpha)/\sqrt{k/n}\}$ for values of x. Then we have $P_{k+1} \leqslant P_k + \varepsilon/n$. From this recurrence relation we find that

$$U_n(x) - \phi(x+2\alpha) \leqslant P_n \leqslant P_k + \varepsilon \text{ i.e, } U_n(x) - \phi(x) \leqslant 2\alpha/\sqrt{2\pi} + P_k + \varepsilon \qquad .. \quad (12)$$

We shall now see that P_k can be made small by a suitable choice of k and n and establish a lemma similar to the one given by Khintchine taking into consideration the existence of p-th order moments of the distribution function $F_1(x)$.

Lemma Given two random variables with mean zero and absolute pth moment $\leqslant \mu_p$ possessing distribution functions $F(x)$ and $G(x)$, then for $\alpha>0$, $F(x) - G(x+2\alpha) \leqslant \mu_p/\alpha^p$ holds for all values of x

Proof. When $x<-\alpha$, $F(x) \leqslant F(-\alpha) \leqslant \mu_p/\alpha^p$ and the lemma is true. When $x>-\alpha$ we have

$$F(x) - G(x+2\alpha) \leqslant 1 - G(x+2\alpha) \leqslant 1 - G(\alpha) \qquad .. \quad (13)$$

Now $G(\alpha) \geqslant 1 - \mu_p/\alpha^p$, and hence $F(x) \leqslant F(-\alpha) \leqslant \mu_p/\alpha^p$ (12) holds for all values of x. Putting $k/n=\delta$ we get from (12) by the above lemma

$$U_n(x) - \phi(x) \leqslant \frac{2\alpha}{\sqrt{2\pi}} + \varepsilon + \frac{\delta}{\alpha^2} \qquad .. \quad (14)$$

Since the variance of $U_k(x)$ and $\phi(x/\sqrt{\delta})$ are both δ therefore chosing $\alpha=\delta^{1/3}$ we have

$$U_n(x) - \phi(x) \leqslant \left(\frac{2}{\sqrt{2\pi}} + 1 \right) \delta^{\frac{1}{3}} + \varepsilon \qquad .. \quad (15)$$

Now we can prove the opposite inequality by taking $V(x,z) = \phi(x/\sqrt{z}) - \varepsilon z$ and the corresponding values of δ and ε will be the same as above . Thus we may state

$$|U_n(x) - \phi(x)| \leqslant \left(\frac{2}{\sqrt{2\pi}} + 1 \right) \delta^{\frac{1}{3}} + \varepsilon \qquad .. \quad (16)$$

which holds for all x. Now if we choose $\varepsilon \geqslant \operatorname{Max}\left[n\int_{-\infty}^{\infty} \xi^2 |\rho(x, \xi, z)| \, dF(\xi), M_1(\delta) \right]$, the inequality (4) is satisfied. Now

$$\frac{\partial^2 V}{\partial z^2} = \int_{-\infty}^{x} \frac{t^4}{z^2} \cdot \frac{e^{-t^2/2z}}{\sqrt{2\pi z}} \, dt - \int_{\infty}^{x} \frac{t^2}{z^2} \frac{e^{-t^2/2z}}{\sqrt{2\pi z}} \, dt + \int_{-\infty}^{x} \frac{e^{-t^2/2z}}{\sqrt{2\pi z}} \frac{dt}{z^2}$$

$$< \int_{-\infty}^{\infty} \frac{t^4}{z^2} \frac{e^{-t^2/2z}}{\sqrt{2\pi z}} \, dt + \int_{-\infty}^{\infty} \frac{t^2}{z^2} \frac{e^{-t^2/2z}}{\sqrt{2\pi z}} \, dt + \int_{-\infty}^{\infty} \frac{e^{t^2/2z}}{\sqrt{2\pi z}} \frac{dt}{z^2} \quad .. \quad (17)$$

$$< \frac{3}{4} + \frac{1}{z} + \frac{3}{4z^2} < \frac{1}{z^2} \text{ when } z \text{ is small}$$

Thus we have $M_1(\delta) = \underset{z>\delta,}{\operatorname{Max}} \partial^2 V / \partial z^2 < 1/\delta^2$ when δ is small. Let us now consider the integral,

$$n\int_{-\infty}^{\infty} \xi^2 |\rho(x,\xi,z)| \, dF(\xi) = n \int_{|\xi|<\tau} \xi^2 |\rho(x,\xi,z)| \, dF(\xi) + n\int_{|\xi|>\tau} \xi^2 |\rho(x,\xi,z)| \, dF(\xi) \quad .. \quad (18)$$

$$\leqslant n\,|\rho(x, \theta\tau, z)| \int_{|\xi|<\tau} \xi^2 \, dF(\xi) + n \underset{z>\delta}{\operatorname{Max}} |\rho(x,\xi,z)| \int_{|\xi|>\tau} \xi^2 \, dF(\xi)$$

where $0<\theta<1$ because of the continuity of $|\rho(x, \xi, z)|$ Now

$$\rho(x,\theta\tau,z) = \left(\frac{\partial^2 V}{\partial x^2}\right)_{x+\theta\tau,z} - \left(\frac{\partial^2 V}{\partial x^2}\right)_{x,z} = \theta\tau \left(\frac{\partial^3 V}{\partial x^3}\right)_{x-\theta'\tau,z} \qquad 0<\theta'<1$$

$$\frac{\partial^3 V}{\partial x^3} = \frac{(x^2/z)-1}{\sqrt{z\pi}} \cdot \frac{e^{-x^2/2z}}{z^{3/2}} \quad .. \quad (19)$$

Thus the maximum value of $|\partial^3 v/\partial x^3|$ when $z>\delta$ is found to be $1/\sqrt{2\pi}\,\delta^{3/2}$. On the other hand $\partial^2 V/\partial x^2 = -x/\sqrt{2\pi z^3} \,.\, \exp(-x^2/2z)$. Therefore $\underset{z>\delta}{\operatorname{Max}} \frac{\partial^2 V}{\partial z^2} < \frac{1}{\sqrt{2\pi\delta}} e^{-1/2}$ which gives

$$.. \quad (20)$$

$$\underset{z>\delta}{\operatorname{Max}} \left| \rho(x,\xi,z) \right| < \frac{x e^{-\frac{1}{2}}}{\sqrt{2\pi\delta}} \quad .. \quad (21)$$

Consider now the integral, $\quad n\int_{|\xi|>\tau} \xi^2 \, dF(\xi) = \int_{|\xi|>n\tau} \xi^2 \, dF_1(\xi)$

If $\sqrt{n}\tau \rightarrow \infty$ then from the existence of p-th order absolute moment of $F_1(\xi)$ $\quad .. \quad (22)$

$$\int_{|\xi|>\tau\sqrt{n}} \xi^p \, dF_1(\xi) < \mu_p = \int_{-\infty}^{\infty} \xi^p \, dF_1(\xi) \quad .. \quad (23)$$

from which we have $\quad \int_{|\xi|>\sqrt{n}\tau} \xi^2 \, dF_1(\xi) < \frac{\mu_p}{(\tau\sqrt{n})^{p-2}} \quad .. \quad (24)$

Thus finally we have $\quad \varepsilon \geqslant \frac{\tau}{\sqrt{2\pi}\,\delta^{3/2}} + \frac{2e^{-1/2}}{\sqrt{2\pi\delta}} \frac{\mu_p}{(\tau\sqrt{n})^{p-2}}$

and $\varepsilon \geqslant 1/n\delta^2$ Thus chosing ε as above we have

$$\left| U_n(x) - \phi(x) \right| < \left(\frac{2}{\sqrt{2\pi}} + 1 \right) \delta^{1/3} + \frac{\tau}{\sqrt{2\pi}\,\delta^{3\cdot 2}} + \sqrt{\frac{2}{\pi e}} \cdot \frac{\mu_p}{(\tau/\sqrt{n})^{p-2}}$$

whichever is greater $\quad .. \quad (25)$

or

$$\left| U_n(x) - \phi(x) \right| < \frac{1}{n\delta^2}$$

We shall now chose δ and τ as functions of n i.e., $\delta=1/n^{\alpha}$, $\tau=n\beta$ in such a manner that the expression on the right has a minimum order in n.

Let $i_1=-\alpha/3$. $i_2=2\alpha-1$, $i_3=3\alpha/2-\beta$, $i_4=\beta(p-2)+\alpha-(p-2)/2$ denote the indices of n in the four quantities on the right. Then the problem is to minimise $m=Max\ (i_1,i_2,i_3,i_4)$ for values of α and β. It is clear that for an extremum, three of four indices i_1,i_2,i_3,i_4 should be equal or they should be equal in pairs. Among seven such sets of extrema we have to pick one that gives an absolute minimum for m. The different sets of solutions are given below :

(1) $i_1=i_2=i_3$

$$\alpha=\frac{3}{7},\ \beta=\frac{11}{14}\left\{-\frac{1}{7},-\frac{1}{7},-\frac{1}{7},\frac{2p-1}{7}\right\}$$

$m_1(p)=(2p-1)/7$ which is inadmissible as $m_1(p)>0$

(2) $i_1=i_2=i_4$

$$\alpha=\frac{2}{7},\ \beta=\frac{1}{2}-\frac{4}{7(p-2)}\left\{-\frac{1}{7},-\frac{1}{7},\frac{p+2}{7(p-2)},-\frac{1}{7}\right\}$$

$m_2(p)=(p+2)/7(p-2)$ which is also inadmissible

(3) $i_1=i_2$. $i_3=i_4$

$$\alpha=\frac{3}{7},\ \beta=\frac{7p-11}{14(p-1)}\left\{-\frac{1}{7},-\frac{1}{7},\frac{p+1}{7(p-1)},\frac{p+1}{7(p-1)}\right\}$$

$m_3(p)=(p+1)/7(p-1)$ which is inadmissible

(4) $i_1=i_{33}\,i_2=i_4$

$$\alpha=\frac{3(p-4)}{11p-21},\ \beta=\frac{1}{2}+\frac{\alpha-1}{p-2}\left\{-\frac{4-p}{28-11p},\ \frac{5p-4}{28-11p},-\frac{4-p}{28-11p},\frac{5p-4}{28-11p}\right\}$$

This is inadmissible for $2<p<4$

(5) $i_3=i_3=i_4$

$$\alpha=\frac{3(p-2)}{11p-14},\ \beta=\frac{11(p-2)}{22p-28}\left\{-\frac{p-2}{11p-14},-\frac{5p-2}{11p-14},-\frac{p-2}{11p-14},-\frac{p-2}{11p-14}\right\}$$

$m_5(p)=-(p-2)\,/\,(11p-14)$ which is admissible for all p

(6) $i_1=i_4$. $i_2=i_3$

$$\alpha=\frac{3(p-2)}{3p-14},\ \beta=1-\frac{\alpha}{2},\ \left\{-\frac{p-2}{3p-14},\frac{3p+2}{3p-14},\frac{3p+2}{3p-14},-\frac{p-2}{3p-14}\right\}$$

This is inadmissible

(7) $i_3=i_3=i_4$

$$\alpha=1,\ \ \beta=\frac{1}{2}\left\{-\frac{1}{3},1,1,1\right\}$$

$m_7(p)=1$ which is inadmissible

Thus we can write

$$|\phi(x)-U_n(x)|<(3/\sqrt{2\pi}+1)\,n^{-(p-2)/(11p-14)}+\sqrt{2\pi e}\ \ \mu_p\,n^{-(5p-2)/(11p-14)} \qquad ..\ (26)$$

We see, therefore, that when absolute moments upto the pth order exist the order of approximation depends upon p. For a given n therefore we can get the best numerical approximation by suitably chosing $\leqslant$ the highest order moment that exists. However since the index of $1/n$ is large for the second term for large values of n the first factor dominates. It is evident however that the above approximation is useless for $p\geqslant3$ for in that case a much better order of approximation has been shown to exist by Liapounof and Cramer. The above approximation is interesting only when $2<p<3$.

Thanks are due to S. N. Roy of the Statistical Laboratory, Calcutta, for help and advice in the preparation of this note.

REFERENCES

1. CRAMER (1937) Random variables and probability distribution, Cambridge.
2. KHINTCHINE (1933) Asymptotische gesetze der wahrscheinlichkeitsrechnung Berlin.
3. LIAPOUNOFF (1900) Sur une proposition de la theorie des probabilities : *Bull. acad, Sci, St. Petersbourg* (5) 13 p. 359-386.

Paper received 25 August 1944.

A NOTE ON CONCENTRATION OF INCOME IN BENGAL VILLAGES

By RAMKRISHNA MUKHERJEA and MONI MOHAN MUKHERJEE
Statistical Laboratory, Calcutta

If $\phi(x)$ stands for number of persons possessing income between x and $x+dx$ rupees, N total number of persons in the community and A and B lowest and highest possible income, then the number of individuals possessing an income of x rupees or more is given by

$$y(x)=\int_{X}^{B}\varphi(x)\,dx \qquad .. \quad (1)$$

Pareto has empirically shown that this can be approximated very closely by

$$y(x)=ax^{-\nu} \qquad .. \quad (2)$$

where ν is always near about 1.5.

If p_x stand for the percentage of individuals possessing x or more and if q_x is the percentage of total income possessed by this percentage of the population, it can be easily shown that

$$q_x=1-(1-p_x)^{1/\delta} \qquad .. \quad (3)$$

where

$$\delta=\nu/(\nu-1)$$

Equation (3) when plotted gives a clear picture of the distribution of income in a community. Gini's concentration ratio ρ wichis given by

$$\rho=1/(2\nu-1) \qquad .. \quad (4)$$

gives a numerical measure of the inequality of income distribution

The object of the present note is to investigate how far the Pareto graduation is applicable to family income in Bengal villages. Two series of material have been used for this purpose. The first relates to the income of 680 families collected in the course of a complete survey of 12 villages surrounding Sriniketan (Visvabharati Institute of Rural Reconstruction) conducted by the Institute in 1937 under the technical guidance of Professor P. C. Mahalanobis. The second series relates to 232 families in six villages (Silimpur, Hatsahar, Amanipara, Ghorsal, Naopara, Roair.) in Bogra District collected in 1941 by Ramkrishna Mukherjea.

TABLE 1. ACCUMULATED FREQUENCY OF FAMILIES.

range of percapita in come in Rs.	Sriniketan observed	Sriniketan graduated	Bogra observed	Bogra graduated
(1)	(2)	(3)	(4)	(5)
10 and above	—	—	232	756
20 ,, ,,	599	907	222	244
30 ,, ,,	469	497	186	126
40 ,, ,,	368	324	127	79
50 ,, ,,	283	233	83	55
60 ,, ,,	213*	170*	56	41
75 ,, ,,	159	128	42	28
100 ,, ,,	99	83	24	18
150 ,, ,,	49	46	10	9
200 ,, ,,	26	30	8	6
300 ,, ,,	15	16	2	3
400 ,, ,,	11	11	1	2
500 ,, ,,	8	8	1	1
600 ,, ,,	7	6	—	—
700 ,, ,,	4	5	—	—
800 ,, ,,	3	4	—	—
900 ,, ,,	3	3	—	—
1000 ,, ,,	3	3	—	—

* The values have been obtained by graduation.

The following equations give the results of income graduation :

$$\text{Sriniketan : } \log y = 4\cdot8882 - 1\cdot4839 \log x \quad .. \quad (5)$$

$$\text{Bogra : } \log y = 4\cdot0077 - 1\cdot6224 \log x \quad .. \quad (6)$$

The following table gives the values of ν, ρ and δ for the two localities surveyed.

locality	ν	ρ	δ
Sriniketan :	1·4839	0·5082	3·0665
Bogra :	1·6294	0·4427	2·5888

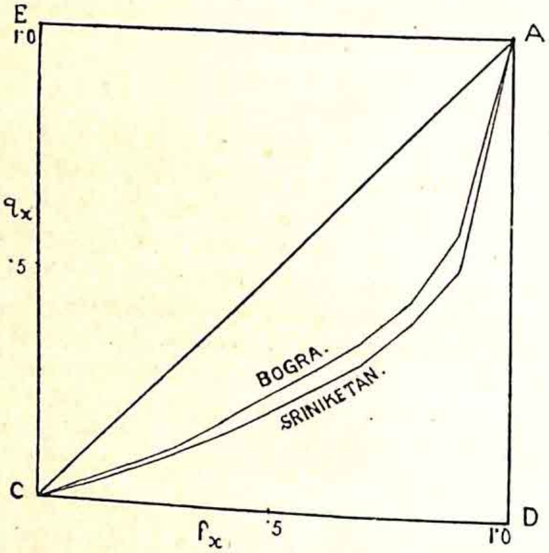
The curves of concentration are drawn from the following equations

$$\text{Sriniketan : } q_x = 1 - (1-p_x)^{0\cdot3261} \quad .. \quad (7)$$

$$\text{Bogra : } q_x = 1 - (1-p_x)^{0\cdot3863} \quad .. \quad (8)$$

and the graphs are shown below.

Chart 1. The concentration curves for Bogra and Sriniketan



The samples are not large enough to draw final conclusions but the close agreement between distributions relating to two sets of villages lying wide apart (one sample from West and the other from North Bengal) probably indicates a general tendency. In wealthy capitalist countries like U. S. A. where production is mechanised the Pareto curve is known to give a good fit with an average value of ν of $1\cdot56 \pm 0\cdot12$ for the period 1914 to 1919. It is interesting to notice that the same graduation is applicable even to the income distribution in Bengal villages where modes of production are still primitive and predominantly agricultural, and actual incomes are small compared to those in U. S. A.

A large value of ρ (which is the ratio of area between AC and the concentration curve to the area of ADC) indicates a markedly unequal income distribution. The position in Bengal villages in this regard appears to be much the same as in most highly industrialized countries. This suggests that the characteristic income distribution (given by the Pareto curve) is more fundamental than methods of production, and that class structure of society is possible in a predominently agricultural and industrially backward society in which mechanical methods of production have scarcely intruded.

The authors propose to discuss the class structure in Bengal village society in a future communication.

Paper received 15 August 1945.

A CRITERION FOR TESTING WHETHER TWO SAMPLES HAVE COME FROM THE SAME POPULATION WITHOUT ASSUMING THE NATURE OF THE POPULATION

By K. K. MATHEN

All-India Institute of Hygiene and Public Health.

In order to test the effectiveness of a medicine for curing a disease, it is tried on a certain number (m) of persons, suffering from the disease in question and for control, another set of persons (n) suffering from the same disease, are not given the treatment. Given the distribution of the duration of time for recovery for the people in the two categories, how can it be tested whether the medicine was effective or not. The standard method, is to calculate the mean duration for recovery of the two samples and apply Fisher's 't' test for the significance of the difference between the two mean values. In applying this test, it is assumed that the samples are taken from a normal population. Another test has been developed by K. R. Nair (1940) for a similar situation in which the median values of the two samples are considered instead of the means. Wald and Wolfowitz (1940) have devised a criterion for testing whether two samples are from the same population based on the distribution of 'runs' when the observations are arranged according to numerical magnitude, no assumption being made regarding the character of the population. The following is an attempt to devise a test, in the above situation, without making any assumption regarding the nature of the population from which the samples have been drawn.

Corresponding to the m persons on whom the treatment was applied and the n persons on whom it was not tried, we have m+n intervals of time representing the duration of time for recovery. Consider these m+n points on a straight line. If the treatment is effective, the intervals of time of the first set of m quantities must be less than the second set of n quantities. Now let us find out how many of the second set of n quantities are being exceeded by each of the first set. The sum of such quantities for all the members of the second set, we use as our statistic (s) for testing whether the treatment is effective. The distribution of this statistic can be found from the following considerations without making any assumption about the population from which the samples have arisen. The null hypothesis, in this case, is that the m+n quantities belonging to the two samples are quantities selected at random from the same population. Consider the n quantities of the control group. They divide the linear continuum into n+1 intervals. Each one of the m quantities of the first sample can lie in any one of these n+1 intervals, with equal probability, since both samples have been assumed to have arisen from the same population. Hence, the distribution of the statistic defined above are the coefficients of x^r in the expansion of the multimomial $(1+x+x+\dots x^n).^m$ For example if m=4 and n=5 the distribution of the statistic is given in the following table.

s	0	1	2	3	4	5	6	7	8	9	10
f(s)	1	4	10	20	35	56	80	108	125	140	146
s	11	12	13	14	15	16	17	18	19	120	
f(s)	140	125	104	80	56	35	20	10	4	1	

From this it is clear that, if 's' is less than 4 the hypothesis tested is rejected at 5 percent level and if 's' is less than 6, it is rejected at 10 percent level.

Probability tables for this distribution for various values of m and n are being prepared at the All India Institute of Hygiene and Public Health Laboratory.

REFERENCES

1. NAIR K. R. The Median in tests by randomisation, *Sankhya* **5**. 4, p. 543-550.
2. WALD, A. AND WOLFOWITZ, J. On a test whether two samples are from the same population, *The Ann. of Math. Stat.*, V. XI., p. 147.

Paper received 6 February, 1946.

PRICE OF AGRICULTURAL COMMODITIES AND LAND TRANSFER IN BENGAL

By N. R. DASGUPTA
Nadia

Bengal is mainly an agricultural province. The conditions of the people are generally very pitiable, so much so that they can hardly afford two humble meals a day sufficient to fill their belly, not to speak of hygienically balanced diets.

People who are a little better off are only a handful in number. And it is only these privileged few who can afford to purchase lands for cultivation.

It will be an interesting study to review the position of the agriculturists in relation to the transfer of lands from one hand to another. With the rise in the price of agricultural commodities, the price of lands for cultivation also goes up. But the increased price of agricultural commodities also brings for the agriculturists an increased revenue so that they can now invest their increased purchasing power in purchasing lands. Rise or fall in the prices of agricultural commodities is thus seen to be very intimately connected with the rise or fall in the number of land-registrations in a particular year. During the period from 1934 to 1943, for example, the latter part of which saw steep increase in the prices of agricultural commodities, the number of land registrations showed a regular tendency to go up counting ten lacs in 1934 and 29 lacs in 1943. For this period, the number of persons in whose favour lands were transferred varied from 30 to 90 per document.

Paddy and jute are the principal agricultural produces of Bengal and it is the one with the higher price that has the main part in controlling the number of land registrations. From 1933 upto 1943, jute-prices were lower than paddy-prices and consequently the paddy prices controlled the registration during this period. Paddy, especially the aman variety, which is the principal variety of the province, is harvested in December and January, and therefore, a study of the December prices of rice with the number of registration of the next year will show the nature of correlation between land transfers and the prices of rice.

Harvest-prices of rice for the 3rd week of December may be obtained for all the districts of Bengal from the Calcutta Gazette of the week. The same can be obtained for all the years under review between 1933 and 1942. The provincial price for a particular year can be easily computed as the arithmetic mean of the prices in all the districts. These, however, it should be noted, give up not prices of a maund or any unit quantity of rice, but the number seers' of rice available for a rupee.

The rise of the number of registrations, however, corresponds to the rise in prices of the previous year. For calculating the correlation coefficient, the series of registrations should therefore be given a lead of one year.

The calculation is easy and gives 9.5 as the mean number of seers of rice per rupee, 1469990 as the mean number of registrations for the period during 1933-43, and the correlation between the two as −.96. The correlation is high and negative and shows that as the number of seers available per rupee decreases i.e. as price increases the number of registrations of cards in the next year floats up. The exact mechanism and interrelations of factors by virtue of which such a phenomenon takes place is yet to be investigated, pending such an investigation the indications of the meagre data as we have, should be taken in their proper light.

Paper received 12 December, 1945.

BOOK REVIEWS

[*EDITORIAL NOTE—We must apologise for the dislocation in Book Reviews caused by the war situation in North East India. Early in 1942 the main library and large sections of the Statistical Laboratory were removed, as an evacuation measure, to Giridih, a place in the interior of the country more than 200 miles from Calcutta. Another difficulty was shortage of paper which led to the introduction of the paper control order by the Government of India in terms of which the paper quota of Journals was reduced to 30 per cent of the normal size. Conditions, however, have now improved materially, and we hope to be able to publish regular Book Reviews in future issues.*]

The Population Problem in India: A Regional Approach by N. V. Sovani. Published by the Gokhale Institute of Politics and Economics, Poona, 1942. Publication No. 8 Pp. XVI+221. Price Rs. 4/- Eight shillings.

The problem of determining the optimum size and structure of the population which a geographical unit can support and the policy which will ensure progress towards optimum conditions naturally depend on the economic resources and environmental conditions in that region as well as on the nature of the population concerned. In a country like India with its vast area and variety in respect both of the human material and environmental conditions, a sound approach would certainly be first to consider homogeneous natural units separately, and then to build up the general picture from a study of such units. In this book Mr. Sovani has adopted this procedure, and has attempted to make a regional analysis of the population problem in India. He has compiled the relevant facts and figures from census reports, agricultural publications and other sources, and has provided a good deal of interesting commentary.

The book is divided into five chapters. The first chapter, entitled 'Framing of the Problem' sets forth the opinions and theories held by a large number of European and Indian writers on population. Twenty-two regions are selected for special study, covering in all about 28% of the area and 50% of the population of India. The rest of the regions are omitted as conditions pertaining to them are considered to be similar to one or other of the selected regions.

The second chapter on the political arithmetic of the regions occupies 126 pages of the book and gives, region by region, the area and total census population in 1931, provisional estimates of population for 1941, a brief description of climatic conditions, the area under principal crops, the density of population at successive censuses from 1881 to 1941, migratory movements of population, distribution of labour in different industries and also a discussion of the age distribution and sex ratio of the population. The chapter ends with four tables in which figures relating to the distribution of population in the different regions are collected together.

In chapter III an account is given of the movement of population in the different regions as recorded in census reports. In chapters IV and V the regions are classified into four broad groups and the characteristic features of each group are pointed out. Chapter VI sets forth certain general considerations and concludes with a plea for a wise policy of conscious control of reproduction.

N. T. MATHEW.

"Student's" Collected papers Edited by E. S. Pearson and John Wishart with a Foreword by Launce McMullen. Published by the Biometrika Office, University College, London, 1942. Pp. 224, Price Fifteen shillings.

It is a service of lasting value to all serious students of statistics that the Biometrika authorities have rendered by re-issuing in a single volume the scientific papers which Mr. William Sealy Gosset (1876-1937) had published between 1907 and 1937 under the pseudonym of "Student". As the editors themselves remark in the preface, 'the central theme of Student's contributions was the application of statistical methods in the research and routine problems of both industry and agriculture' and "he was a pioneer worker in a field which, during his later years, was rapidly expanding, and his work is intimately related to the historical development of his subject". Of his twentyone odd papers scattered over a number of journals, no less than fourteen appeared in Biometrika and are probably familiar to readers of that journal. But papers of historical importance such as 'Yield Trials' in Bailliere's Encyclopaedia of Scientific Agriculture (1931) would not have been easily accessible to the majority of statistical workers except through the present commemoration publication undertaken by Professor Pearson and Dr. Wishart.

It is impossible in this brief review to attempt any evaluation of "Student's" work. Suffice it to say, he may be called the father of the 'small sample theory' and the originator of what has come to be known as 'Studentised statistics'. In the design of field experiments in which Professor R. A. Fisher introduced revolutionary techniques, Mr. Gosset made notable contributions by way of expository articles, though the proviso of randomisation required in the Fisherian technique remained a bone of contention between them. The masterly way in which he led the defence for balanced arrangements and in particular for Beaven's halfdrill strip method may confound the learned judges of posterity and more so because death removed the counsel before he could conclude the arguments.

The book has two additional attractions of a personal nature in a photograph of Mr. Gosset and in the foreward by Mr. McMullen, one of Mr. Gosset's Colleagues in the Dublin Brewery.

K. R. N.

Table of Powers initiated by J. W. L. Glaisher, extended by W. C. Bickby, C. E. Gwyther, J. C. P. Miller and E. J. Jernouth. British Association Mathematical Tables Vol. IX. Published by the University Press, Cambridge, 1940. Pp. 132. Price Rs. 22/8/-.

This is the ninth publication of the very useful series "British Association Mathematical Tables". The present volume is based on tables computed under the supervision of J. W. L. Glaisher about seventy years ago, and later extended by W. G. Bickby and others, and contains the following tables :—

(1) Upto 50th power of numbers from 1 to 99 ; (2) Upto 16th power of numbers from 100-299 ; (3) Upto 12th power of numbers from 300-1099 ; (4) Tables for evaluating sum of powers up to 50 ; (5) Values of Binomial Coefficients up to $n = 50$.

There is an introduction by Mr. J. C. P. Miller and a bibliography of important mathematical tables which contains sets of complete integral powers. The present volume, like others in the same series, will be extremely useful for computational purposes.

P. K. BOSE.

A Review of Price control in India by S. G. Beri, Professor of Economics, Sydenham College of Commerce & Economics. Published by the New Book Company, Bombay. Price Rs. 1/4/-.

The publication deals with one of the most intricate and vital problems of wartime living. The argument is closely reasoned and the lucid style enables even a man without any technical knowledge of economics to grasp the subject matter thoroughly. The author bases his conclusions on statistics relating to cost of living index, index number of prices, and currency circulation. The index figures are known to be inadequate and out of date in many ways, but probably supply a sufficiently reliable indication of general tendencies.

Price is determined by the equilibrium of the forces of demand and supply. Having regard to the exigencies of war economy, which is characterised by increased demand in relation to supply and inflationary tendencies arising from a stupendous increase in government expenditure and therefore in the purchasing power of the people as a whole, price regulation by itself is not adequate to keep down prices and stabilise the cost of living. Supplementary measures, such as introduction of enhanced rates of taxation, rationing of scarce goods, imposition of restriction on export and import, stabilisation of remuneration received by individual workers, stabilisation of agricultural prices, diversion of earnings of individuals from the purchase of unessential articles into saving, and discouragement of credit and instalment buying must be enforced to ensure success in the price control regulation. The author has discussed the various techniques and practices of price control adopted by countries like Germany, Canada, U.S.A. and U. K.

Apart from difficulties of price control experienced in other countries, India has inherent difficulties, like the vast size of the country, the enormous rural population, division of the country into several separate administration (provincial and state) units, lack of organized consumer's opinion, opposition of business interest to price control, disturbed political situation, large purchases made by the Government of India on behalf of the British Government and partly for India's own defence requirements and to some extent for reciprocal aid to U.S.A. forces stationed in the country.

To safeguard against such additional difficulties Government should have adopted more stringent measures. The author has rightly observed that "the large expansion of currency could have been partially avoided by importing goods such as machinery and other accessories of industrial development from or through the British Government or by raising British loans in India. The inflationary rise in price could have been checked by initiating a vigorous production drive in the country and ensuring fuller employment of the factors of production for which there is great scope in the country. Absence of unified and uniform price-control by central government, the piecemeal and unilateral control imposed by the provincial and state governments, absence of a coordinated survey of food requirement of the different regions, inadequate administrative machinery for implementing price control measures are some of the deficiencies of the price control policy in India."

The author is not in favour of withdrawal of price control. He is of opinion that more direct participation by Government in the actual procurement and distribution of supplies would be the proper course.

T. CHOUDHURY.

Savings & Investments in India by A. K. Sur, M.A., with a foreword by Mr. Nalini R. Sarkar. Published by the Indian Economist Ltd., Calcutta, 1944. Pp. 34. Price -/12/-.

This study of trends in different forms of investment in India is specially welcome at a time when large industrial developments are expected to occur in India. The author's finding is that owing to a variety of factors operating together in the thirties of this century indigenous capital showed symptoms of overcoming its shyness. Forms of investment other than real estates and jewellery, that is, both commercial and industrial investments, are finding increasing favour with Indian investors. With surplus funds created during this war going mostly into Government loans as well as life insurance policies and new industrial floatations, this change towards modernisation of investments may be expected to continue if the fiscal policy also does its part for the development and 'protection' of Indian industries. To banish the 'liquidity-preference-complex' from the mind of Indian capital owners the author suggests the establishment of a well-organised and broad-based investment market offering easy facilities.

Mr. Sur deserves much credit for basing his statements on statistics. It is a great pity, however, that he has not given references to sources of information. One obvious error has crept in at page 5 where he is quoting the statistics of the income-tax payers in 1938 to show the high dispersion coefficient of the income distribution of this country: the frequency in the income-group (Rs. 1000-1,999) is stated to be only 500 which is obviously due to the exemption of the important categories of incomes falling in that group from April, 1936.

DURGADAS BASU.

Conflict of Economic Ideologies in India: An attempt at Reconcilliation by Sudhir Sen B.Sc., (Econ.), Ph.D. Publication No. 1 of Visva-Bharati economic research studies. Published by Visva-Bharati, Santiniketan, 1941. Pp. 44. Price -/8/- as.

The conflict between the charka and machine school of thought is considered here from the point of view of general principles, and it is finally concluded that at the present stage of economic development, India can ill afford to switch over entirely to large scale industries. On the other hand, blind faith in small scale industries alone is also detrimental to progress. What is desirable is to hit a golden mean of these two types; Warthemberg's countryside industries are cited as a suitable form.

As already stated, the author deals with broad principles and general theories. This is the usual line of economic discussions in India; facts are presented in support of theories which are believed on *a priori* grounds. Such an approach is not likely to be of much practical value. What is needed is a patient and objective study of real situations based on relevant facts and figures. Until this is done economic research is bound to remain abstract and academic. It is not suggested that Dr. Sen's conclusions are wrong. They may be perfectly correct but what is needed is establishing their validity on objective grounds. The scope for village study at Sriniketan is tremendous. If this is fully utilised this centre should be able to make significant contributions to the economics of rural reconstruction.

M. MUKHERJEE

Examinations in India by D. N. Mukerji. Published by Hind Kitabs, Bombay, 1944. Pp. 46 Price -/12/- as.

Our dissatisfaction with the current system of examinations is today so complete and deep-rooted, that we are no longer conscious of it. Advocators of reform are no longer received by the authorities even with indifference; they are simply ignored. Yet criticisms of the system crop up from time to time with an odd regularity—like hardy annuals.

Unfortunately, however, criticisms levelled against examination authorities are most often too fragmentary and unsystematic to be of much use. This book has the great merit of going into at least a scientific, if sketchy, analysis of the subject of examinations. The time is now ripe for launching an organised campaign against the appalling foolishness of our present system. The Progressive club of the Allahabad University which sponsored the publication of this book must be thanked for taking the first step towards this.

The author regrets that the Sargent scheme should leave out the problem of examinations altogether, but he stands ready with long excerpts from the Indian Universities Commission Report (1902), the Calcutta University Commission Report (1917-19), the Zakir Hussain Report(1938), to show that official indictments of the current system have not been lacking. There are three things, he tells us, which an examiner ought to know—the purpose of examinations, the relationship between teaching and examining, the distinction between knowledge and skill. But these are precisely the three things in which the modern examiner's ignorance is most manifest.

Statistics, in India, is such a young and undervalued science that it is particularly heartening at this point to see the author recommend the use of statistics to solve the problems of examinations. To reinforce this point, he gives a long excerpt from 'Marks of Examiners' by Hartog and Rhodes. "Every Examiner," it runs, "should acquire as part of his training, some knowledge of the elementary statistical principles involved in such work........the art of examining......rests on scientific principles and involves scientific technique ; and this technique must, to a large extent, be statistical . Even so simple a task as combining marks for different questions cannot be rightly understood without a knowledge of what must occur when marks, more or less uncorrelated, are averaged or summed. Nor can a proper allocation of marks be made without some explicit theory as to the curve of distribution which marks should approximately follow" etc. A survey of the marks gained by about a thousand High School Examination candidates for each of the years 1928, 1929 and 1930 is here summarised into a table which provides the basis for the conclusion that pass and failure are determined at present by a level which is, to say the least, arbitrary. "Actual results", the author goes on to say "as 'passes' and 'failures' should be decided on the basis of the achievements of the average group, i.e. in the light of the (Mean I Standard deviation) figures".

Coming to the chapter on suggestions one is reassured that this is not one of those over-enthusiastic and over-theoretical criticisms which are content to draw the line at mere indictment. The suggestions are built up into a coherent system which should be capable of application. One might allow himself to hope for a moment that such attacks on the professional and scientific front will at last begin to undermine the self-complacence of our Universities. But the state of education in India being what it is, cynicism, if not inevitable, is at least natural enough.

C. Das Gupta.

Building the Canadian West by Professor James B. Hedges. Published by Macmillan & Co., New York, 1939. Pp. 421.

The main theme of this well-documented book is the achievement of a private syndicate, the Canadian Pacific Railway, in colonizing the prairies of West Canada. In building the railways in Canada, the syndicate and the Dominion Government were not interested only in a "quit-cash transaction" ; fundamentally they wanted permanent settlers and communities which would ultimately create traffic. Preliminary investment was therefore heavy, and the terms between the two parties are described precisely in the following quotation :

> "In October 1880, the syndicate signed a contract with the Dominion government to build two thousand miles of railway westward from the Great Lakes to the Pacific coast. The syndicate received a cash subsidy of 25 million dollars and a land grant of 25 million acres. Granting the land was one thing, locating and selecting it was another. Here Canadians profited by the experience of the United States. In the latter country, railways had received a stipulated amount of land per mile of line built ; and the land was contiguous to the railway. The C.P.R. on the other hand, received a fixed amount of land, 25 million acres, regardless of the length of its line ; moreover the land did not need to be contiguous to the railway. It was here that the American experience of some of the syndicate members served them well. The contract included a "fairly fit for settlement clause" which meant that the syndicate could choose their 25 million acres critically and choose it anywhere. Actually, the land was selected in what are now the three prairie provinces of Manitoba, Saskatchewan, and Alberta."

The history of development starts from 1881 and has been traced till 1937. The Railway had to pass hrough periods of depression ; crops failed, weather revolted against settlement, the war of 1914-18 brought a recession in sales and in its wake an economic depression. The authorities of the Railway launched irrigation projects, introduced better types of wheat, sponsored a Department of Natural Resources and operated an ocean service for immigrants across the Atlantic. In regions of inferior or more inaccessible lands, the company embarked on a programme of assisted and supervised settlement e.g. the ready-made farm, the loan farm idea. "The policy was not philanthropy but enlightened selfishness on the part of the company".

By means of extensive advertising with centres in Europe, there was a great tide of immigration and no selection was made during the rush. The chapter on 'Advertising the West' describing the technique and the organization is illuminating for advertising agencies even of today. The policy of land settlement till 1914 had been entirely defined by the company. In post-war years the problem of demobilized soldiers inspired the Government to encourage a selective immigration policy and the latter adopted a directive attitude towards the company. The Dominion Government started 'training farms' with subsidy from provincial governments and its own exchequer. British settlement was preferred naturally by the Canadian Government.

All the thirteen chapters of the book bear the stamp of thorough and able ransacking of the carefully preserved archives of the Canadian Pacific Railway. The list of reports and memoranda exchanged between the Company and the Government convinces the reader of the vision, energy and hardheadedness of the Railway Company : the role of the Government was possibly to take advantage of 'capitalist expansion'.

The author has not discussed social aspects of the scheme while describing its development. This explains the pessimistic note of his conclusion. "Speed was sometimes unwise and made for waste of national resources, human and material". Absence of a study of the new economic and social organisation that sprang up in the way of the colonisation has made the author think that "the Company's organisation......could do little more than mark time". The expansion was planned for agricultural production without keeping in mind the need of equating it with other human demands. Price depression has been the result, and certain settlements have also proved costly due to frequent droughts. The incentive of this war has fostered the necessary development of industries. One would have liked to have some idea of the wider social and economic developments which had occurred in the course of colonization.

The value of the book in relation to conditions in India lies in its able compilation of facts dealing with development of waste land. This history may serve as a guide to the various plans which are being made in India for land development.

P. C. BANDYOPADHYAY

The Drama of Weather by Sir Napier Shaw : Published by Cambridge University Press. 3rd Edition, 1939. Pp. 307 with index. Price 10s. 6d.

Sir Napier Shaw believes that 'weather is the work of an accomplished artist'. He has therefore infused a dramatic background throughout while interpreting the sequence of events, in the sky, and in this drama clouds are the star players. He himself calls the contents of the book as 'unconventional essays'. In the first few pages he conducts the reader to the stage of the sky and introduces the different actors sunshine, wind, temperature, pressure and lastly the clouds. Sir Napier asks us to watch the pageant which has been transforming yesterday into to-morrow. In concluding the prologue—pageantry in the sky—, he says, 'there is perhaps no subject more suitable for illustrating what is meant by science, the gradual progress from the chaos of the first impression, through co-ordination and classification to order'.

While beginning the subject the author looks behind for examples from the remote ages of their interest in weather, and then gradually turns to presenting the accumulated materials which do not show any sign of nearing the end. Instruments which were gradually developed to watch the drama, or rather to clothe the artists for the show, are described with details of their successive improvement. Records of observations which can now be exchanged between and collected speedily from far-flung weather stations by wireless have been presented more in the form of visual images than in dry language. And his heading of the chapter is rightly 'the weather map presents the play'. In fact, the essential merit of the book is the number of beautifully reproduced illustrations. And that is why a reader can never grudge, which he does with many of the popular books of science—, that he has been given some knowledge which has to be mentally retransformed into pictures. The book is really a bold but eminently successful venture in giving an ordinary reader such wealth of information which also serves so admirably as a compendium for the meteorological workers.

In this edition a chapter has been added dealing with the weather data in relation to the growth of crops. Agricultural meteorology has now a literature of its own, which has grown rapidly with the introduction of statistical methods.

'Where our rains come from' is the title of another new chapter which explains the mechanism of wind currents and their energy. The new light of air-mass fronts which has been thrown on meteorology by the Norwegian scientist, Bjerknes, has been given a short focus. It has been possible to develop as early as

1937 a method of forecasting for periods of ten days in advance. Its success has been such as to become a weapon of strategy in warfare, and it is told by newspaper men that Hitler had advance information of King Mud's coup d'etat in Polish Campaign of 1939. Background of the whole process is the understanding of 'broad-weather situation', a term used by the German meteorologist Fanz Baur. "Just as weather connotes a number of processes lasting perhaps half a day or a day, the conception of a broad weather situation includes a series of individual weathers; it is a condition of the atmosphere which controls the weather for several days, remaining sensibly constant in spite of the changes in the latter from day to day. Towards the study of the broad weather conditions, a very early contribution was made by P. C. Mahalanobis, who had shown in 1922 that the relations of pressure-temperature conditions at 4 km were closer than those found by Dines at 9 km. Statistical methods used by Mahalanobis thus located the seat of activity in the upper air at a lower level, and observations here are comparatively easy for collectiong the necessary information for predication

From start to finish Sir Napier has been essentially a poet. But he has sermons also for the rigid and red tape methods of weather broadcast which are 'crown copyright reserved' and therefore hamper progress of research. The book is unlike most others of popular series. It does not develop an awe-inspired veneration in the mind of the reader for the mysteries delved into by scientists, but adds to the scanty scientific knowledge, both analytic and synthetic, of the common man. Art paper on which the book is printed has made it possible to have before the reader the 'scenes' of the drama, and the pictures have mellowed the hard core of scientific knowledge.

P. C. BANDYOPADHYAY

The Forming of Federations R. R. Kale Memorial Lecture by S. G. Vaze, Editor, Dyan-Prakash Published by the Gokhale Institute of Politics and Economics, Poona, 1943. Pp. 32. Price Rs. 1/8/-.

The chief task which Mr. Vaze sets himself is to demonstrate the importance of the principle of self-determination in forming Federations. His whole aim is to make us see that it is not by hoodwinking our differences, but by recognising them that we can approach a basis on which unity is possible. In a country like India, torn by the conflicting interests of various nationalities, cultures, religions, to try to overlook differences in order to overcome them is to travel as far away from the solution as possible. Modern history has innumerable analogies to prove this. In Canada, the effort to establish a unitary state failed, and the authorities were compelled to divide united Canada into the provinces of Ontario and Quebec. The French Canadians and the British Canadians could come to a settlement only on the Federal basis. And even Lenin who stood for "the closest possible amalgamation of large territories" had nothing but Federation to choose. Unity, he found, 'must be attained by means of teaching, propaganda, and voluntary consent'; for unity achieved by coercion contains within itself only the seeds of its own destruction.

This bird's-eye view of western constitutions (Canada, Soviet Union, Finland, Czechoslovakia, Yugoslavia and Pre-Hitler Germany) ought to be enough to convince us that the 'Akhand Hindustan' talk is prompted more by sentimental than political considerations. Where self-determination is the topic, one often hears it said that self-determination is good, but it should not go by race or religion. And it is of course true that these should not enter into politics as active factors ; "But it (race or religion) is a factor capable of immense potentialities for mischief, and the best way of sterilising it to a maximum degree is to take it deliberately into account and make it the basis of administrative sub-divisions". Quoting Selborne in defence of this idea, Mr. Vaze says "where there is a measure of distrust between peoples of different nationalities, the best way to overcome this difficulty is for both races *to avow it* (Italics mine). and at the same time for both to admit that the only possible form of union is one which the other party will accept as well as themselves". One hopes that such reflections on the history of modern constitutions will open our eyes to the situation as it really is and not as we imagine it to be. But to learn from history is not given to every country, and one fears at moments that it is not given to India.

Nearly two-thirds of this booklet is given to the discussion of western, chiefly European constitutions. All this is no doubt very relevant ; but it is relevant only as a historical context in which the Indian problem must be seen if it is to be properly understood. It is not therefore obvious why the author should enter into the concretest details in the discussion of this context and restrict himself to the most general statements with regard to India. The question of self-determination comes up in India chiefly with reference to the demand for Pakistan. Yet not once is the question of Pakistan mentioned or the meaning of Pakistan in the light of self-determination discussed. But was this because Mr. Vaze was speaking in Poona, the citadel of the 'Akhand Hindustanites' ?

C. DAS GUPTA.

SANKHYĀ
THE INDIAN JOURNAL OF STATISTICS
Edited By P. C. MAHALANOBIS

VOL. 7 PART 4 1946

A SAMPLE SURVEY OF AFTER-EFFECTS OF THE BENGAL FAMINE OF 1943

BY P. C. MAHALANOBIS, RAMKRISHNA MUKHERJEA, AND AMBIKA GHOSH
Statistical Laboratory, Calcutta.

PREFACE

The present report is being published with two objects in view :

(1) To give a concrete idea of the economic background of the Bengal famine of 1943 and its after-effects.

(2) To indicate the kind of results which can be obtained by socio-economic enquiries carried out by the method of sample surveys.

The present enquiry had to be carried out under great difficulties caused by lack of adequate financial support. It is hoped, however, that the results discussed in this report would enable Government and the interested public to judge how far such enquiries are likely to be useful in practice.

The senior author has been advocating for some considerable time the organization of a comprehensive multi-purpose sample survey scheme in Bengal which would be in operation throughout the year. Once this is done, it would be possible to collect information not only about acreage, rate of yield and total production of important crops in each season (as is being done at present in the crop survey scheme) but also about various economic and demographic factors such as cattle, cultivation methods and implements, drainage and irrigation, rural unemployment and indebtedness, village crafts and industries, roads and transport, marketing, prices, cost of living, patterns of consumption, literacy, public health, epidemics and diseases, birth and death rates, growth of population, etc, and also, if desired, changes in the national dividend in the rural sector.

A standard programme of work would be set up so that information relating to certain items would be collected every month, or in each crop season; other items once every year or once in two years and so on. The population of the country with sex and age or occupational distributions can be estimated by the sampling method from time to time or every year.

Besides the standard programme, the sample survey organization would be ready at hand to carry out rapid *ad hoc* enquiries at short notice. If there is any sign of failure of crops, or outbreak of an epidemic, or distress caused by a natural calamity like flood or earthquake, the sample survey organization would be able to collect information relating to conditions in affected areas in the course of a month or two. Such an organiszation would make it possible for Government and the general public to maintain almost day to day contact with vital socio-economic conditions in the country. The senior author has already submitted a scheme which is at present under consideration by the Government of Bengal.

GENERAL SUMMARY

History of the survey. Bengal was seriously affected by a famine in 1943. Material relating to the economic life of the rural population was meagre, and reliable information relating to the famine was simply not available. A sample survey was conducted in 1944-45 by the Indian Statistical Institute under the guidance of P. C. Mahalanobis in collaboration with K. P. Chattopadhyay, Head of the Department of Anthropology, Calcutta University, to investigate the after-effects of the famine.

2. Originally it had been intended to carry out the field survey with the help of University and College students during the summer vacation of May, June and July 1944. In order to enable the work being done on an adequate scale the Government of Bengal were approached for financial assistance and a grant-in-aid of Rs. 25,000. was sanctioned. There was however a delay of two months in sanctioning the grant which completely upset the time programme, and the field survey had to be extended from the end of July 1944 to early February 1945. Owing to the setting in of rains and other difficulties, expenses had increased very considerably. The field survey had therefore to be curtailed to some extent for lack of funds, and the analysis of the material was delayed for the same reason.

3. The present report gives certain results based on a statistical analysis of a portion of the data collected during the enquiry but much material still remains unanalysed. Besides the after-effects of the famine the present survey has supplied valuable information relating to economic conditions in rural Bengal.

4. *Occupational groups.* The effect of the famine can be fully appreciated only when considered separately for different socio-economic sectors of the population. In the present report the family is the unit of survey and consists of all persons taking food from the same kitchen (in accordance with the Census definition). Family occupation has been defined as that particular occupation from which the greater part of the total family income is derived

5. Agricultural occupations have been sub-divided into four classes : 'agriculture' consisting of *ryots* (peasant proprietors) who actually cultivate their own land or land taken on share (*barga*) basis but who do not work as hired labourers ; 'agriculture & labour' who not only cultivate their own land or land taken on a share basis but also work as agricultural labourers ; 'agricultural labour' consists of persons who do not own any land or only a negligible area ; and 'non-cultivating owner' is a heterogeneous group consisting of owners of large areas as well as widows and invalids who are obliged to cultivate their own land by hired labourers or on a share basis.

6. *Economic background of the famine.* The number of families in rural Bengal (in 86 subdivisions covering about 70,000 square miles) in 1943 is estimated at 102.4 lakhs (10.2 millions) consisting of about 552 lakhs (55.2 millions) of persons with an average family size of 5.4 persons . The official figure for the total area under cultivation in normal times is about 250 lakhs (25 millions) of acre including double (or sometimes triple) cropped areas. (There are, however, reasons to believe that the actual area under cultivation is appreciably larger than the official figure). Paddy forms the staple food and money crop of the province and constitutes about 88% of the cultivated area. The area under *aman* paddy (the major winter crop) was probably something like 180 or 190 lakhs of acre just before the famine.

7. The average yield in round figures is about 10 maunds (about 820 lbs.) of rice (not in husk) per acre. The average size of the family, defined as persons having food from the same kitchen (in accordance with Census practice), is about 5.4 in rural areas; and the average over-all consumption of rice is estimated at about 4 maunds per head per year. The subsistence level is usually taken as 2 acres of paddy land per family on an average.

8. The actual position in Bengal even before the famine was precarious. About one-third of all rural families did not own any paddy land while two-fifths had less than 2 acres so that about three-fourths of all families had either no paddy land or owned less than 2 acres. The over-all average for the province as a whole was about 1.8 acre of paddy land per rural family which was below the subsistence level or on the border line. It is also worth mentioning that, averaged over a number of years before the war, there was a small but net import of about one per cent of total production of rice and other cereals into the province.

9. The cattle position also was not satisfactory. The sample survey showed that the total number of plough cattle in 1943 was 79 lakhs giving a share of 4½ acres of paddy land per pair of bullock which according to many economists was only just adequate or fell slightly short of requirements.

10. In a majority of subdivisions (26 out of 41 surveyed) more than 70 per cent of the families owned less than 2 acres of paddy land per family, and the average owned per family was something between 1.5 and 2 acres (Table 5.2). In fact classification of subdivisions by amount of paddy land owned per family before the famine was found to be roughly parallel to the degree of incidence of famine conditions. This is, of course, just what is to be expected. Subdivisions in which there were more families with paddy land below subsistence level were naturally more vulnerable to the famine.

11. *Land sale and mortgage.* During the period April 1943 to April 1944, 9.2 lakhs of families sold their paddy land in full or in part out of whom 2.6 lakhs had sold their land in full and had lost their only or chief means of livelihood; 6.7 lakhs of families mortgaged their paddy land of which 1·03 lakhs had also sold a part of their holding. In other words, nearly 16 lakhs of families (about one-fourth of the number who had owned paddy land before the famine) had either sold in full or in part or mortgaged their paddy land during the famine period.

12. Sales in full were most important in the group owning less than 2 acres; 2.4 lakhs out of 40.1 lakhs or 6.1% of families were obliged to do this and lost their chief means of liveli- hood. Only about 20,000 out of 16.2 lakhs or 1.2% of families in the middle group (owning from 2 to 5 acres) sold all their paddy land; and less than 4000 or out of 8.8 lakhs 0.4% of families in the upper group did the same thing.

13. Sales in part were least frequent (about 3 lakhs out of 40.1 lakhs or 7.4%) in the lowest group owning less than 2 acres; most frequent (2.5 lakhs out of 16.2 lakhs or 15.3%) in the middle group owning from 2 to 5 acres; and quite frequent (1.1 lakhs out of 8.8 lakhs or 12.5%) in the upper group owning above 5 acres of paddy land. Families in the lowest stratum who sold paddy land were obliged to do so in full rather than in part. In the upper group, sales were most probably due to the very high price of paddy land which was a characteristic feature during the famine year. In the middle group, sales were probably due either to distress or to the desire to make profit depending on the economic position of the family.

14. The position may also be considered by occupational groups. Mortgaging or selling paddy land in part was heaviest (7.2% and 9.9% respectively), but sales in full were comparatively low (2.4%) in the group 'agriculture' indicating that there was general impoverishment but no pauperization on a large scale among cultivators owning their own land. The group 'agriculture & labour' was much more severely affected (6.0% sales in full, 6.7% of sales in part, and 7.6% of mortgage) showing that not only impoverishment but pauperization was widespread in this sector. The number of families in the group 'non-cultivating owner' is small, only 5.9% of the total ; the majority is well off so that the proportion of families selling or mortgaging paddy land was low. The groups 'agricultural labour' and 'others' had little of paddy land and naturally did not participate much in sales or mortgate.

15. Considering the position by geographical regions it is seen that subdivisions in which families owned, on an average, less paddy land were more severely affected by the famine. Families owning less land were obliged to sell it more heavily which increased further the number in the lower economic levels. Mortgaging was also comparatively heavy among less favourably placed families.

16. The most important point to note is that, during the famine, 2.6 lakhs of families (out of 65 lakhs owning paddy land) had totally lost their holdings and were thus reduced to the rank of landless labour. The famine transferred an appreciable number of families from the group 'agriculture' to 'agriculture & labour' and from 'agriculture & labour' to 'agricultural labour' ; and thus accentuated inequalities in the distribution of paddy land in the province.

17. Another fact is worth noting. Out of the total of 7.1 lakhs of acres of paddy land sold during the famine, only 2.9 lakhs of acre had been purchased back in the villages. Roughly 4.2 lakhs of acres of paddy land had thus passed to outsiders, possibly 'non-cultivating owners' residing in urban areas. In 'agriculture', 'non-cultivating owner' and 'others' groups of families roughly half the land sold had been purchased back. But among 'agriculture & labour' and 'agricultural labour' only about 10 per cent of the land sold have been replaced by purchase showing that net loss of paddy land was most severe in these two classes.

18. *Loss of plough cattle.* As already noted, the total number of plough cattle of Bengal before the famine was just adequate or fell short of requirements for the cultivation of *aman* paddy, the main crop of the province. The net loss of plough cattle was about 10 or 11 lakhs (about 13%) during the famine period which must seriously affect agricultural operations in future. Out of this, 9.4 lakhs (65%) were lost by sale, and 5 lakhs (or 35%) of cattle by death. Only about one-fourth of the loss (3.5 lakhs) was replaced by purchase. About 3 lakhs or 8.5% of families of rural Bengal had probably lost all the cattle they had before the famine making it difficult or practically impossible for them to carry on normal agricultural operations.

19. Sales of cattle largely exceeded purchases showing that transfers had taken place not merely from one rural family to another, but that large purchases had been made by outsiders (possibly by contractors for the supply of meat for army consumption).

20. The average number of plough cattle owned per family naturally decreased in all groups but the reduction was largest in 'agriculture' and 'agriculture & labour, the two groups mainly concerned with cultivation. Even before the famine a considerable number of these families did not own any cattle ; their number increased during the famine and further aggravated the cattle position.

21. *Economic deterioration.* It is possible to get some idea of economic deterioration by the number of families transferred from occupations at a higher economic level to occupations at a lower level. Assessed by such methods, it was found that about 7 lakhs of families in rural Bengal had suffered a lowering of economic status with consequent decrease of earning power during the famine ; adopting 5.4 as the average size of the family the total number of persons whose economic position had deteriorated was thus about 38 lakhs. It must be remembered that in this method deterioration within each occupational group has been ignored, that is, families retaining their occupational classification have not been taken into consideration.

22. In actual numbers, heaviest deterioration had occurred among families which originally belonged to the two occupational groups 'agriculture' and 'agriculture & labour' followed closely by rural 'trade' and 'craft'. Impoverishment had been however proportionately greatest among families following rural 'trade' and village 'crafts'.

23. *Destitution.* The rate of destitution (that is, proportion of persons living on charity) was 1.07% at the time of the 1931 Census ; the corresponding figure for 1941 is not available as relevant tables were not prepared. Assuming that the rate had remained steady, the total number of destitutes (on the basis of a rural population of 55.2 millions) would have been about 6.0 lakhs in January 1943. The sample estimate was about 7.5 lakhs so that at the beginning of 1943 there had been already an increase of 1.5 lakhs of destitute persons. There was a further increase of about 3.3 lakhs of destitutes between January 1943 and May 1944. At the normal rate of 1.07% the total number of destitute families in May 1944 should have been 6.0 lakhs. The sample estimate was 10.8 lakhs showing that 4.8 lakhs of persons had been rendered destitutes under war and famine conditions in Bengal.

24. Compared to the number of destitutes in January 1943 the largest proportion had come during the famine period from younger age groups. The proportion of destitute women was greater than destitute men especially among adults belonging to the age-group 15 to 50 years which created a serious socio-economic problem.

25. In actual numbers, 'agricultural labour' (that is, landless labour) had contributed the largest share of destitutes during the famine ; other groups in order of importance were 'agriculture', 'agriculture & labour', 'craft', 'fishing' and 'trade' ; least affected were 'transport', 'non-cultivating owner' and 'husking paddy'. Proportionately to their total numbers most seriously affected were 'fishing', then 'agricultural labour', 'husking paddy', 'craft' and 'transport''

26. The chief cause of destitution was death of earners, next in importance was sickness or unemployment of earners.

27. It is worth noting that economic deterioration and destitution had affected different occupational groups in different ways. Economic deterioration was more important than destitution among the groups 'agriculture', 'agriculture & labour', while the reverse was the case among 'agricultural labour' and 'others'. Among 'others' again, economic deterioration was relatively more important in rural 'trade', 'transport', 'non-agricultural labour', 'profession & service', and 'craft' while destitution was relatively more important in the groups 'husking paddy' and 'fishing'. Among 'non-cultivating owner' both factors were comparatively small. Evidently occupational groups which could live on their assets were able to resist destitution to some extent but groups in a more precarious position had succumbed more quickly to famine conditions.

28. *Deteriorating conditions before the famine.* Even in the pre-famine period (January 1939 to January 1943) the proportion of families suffering economic deterioration and destitution was much higher than the proportion who had improved their position. About 6.84% of families had suffered economic deterioration against 3.32% who had improved their economic level while the position of 1.07% was not clear. This shows conclusively that economic deterioration had set in definitely in the pre-famine period and that the famine itself was its culmination.

29. *Accelerated changes during the famine.* During the famine period, rates of change became more rapid. Improvement during the famine period was relatively twice as great as that in the pre-famine period ; but this was set off by a three times greater rate of economic deterioration and twelve times greater rate of destitution. The famine period was thus one of accelerated economic changes. Improvement of economic conditions was also quicker but was restricted to a comparatively small number of families. Deterioration and destitution were even more violently accelerated and were shared by a much larger number of families. Roughly 85% of families, however, succeeded in maintaining their status quo showing that a large degree of economic inertia persisted even under famine conditions.

30. *General review.* The general picture is thus clear. Certain areas were very seriously affected ; others to a moderate extent ; and still others only to a slight extent. This patchy nature of the incidence of famine conditions indicates that large regional differences in economic conditions had existed even in normal times, and were further accentuated durin the famine. The poorer sections of the community, especially landless labour, fishermen and village craftsmen, were most severely affected and many were rendered destitutes ; the middle groups who had land of their own and other assets were naturally less vulnerable while the upper stratum remained immune and sometimes even prospered. There is clear evidence to show that economic deterioration on the whole had set in even in the pre-famine period ; a comparatively small number of families were improving their economic position while a far larger number were suffering from economic deterioration or destitution. During the famine period (January 1943 to May 1944) the whole process was greatly accelerated but the general nature of changes remained much the same. A small number of families improved their position while a much larger number were impoverished or rendered destitute. The famine of 1943 was thus not an accident like an earthquake but the culmination of economic changes which were going on even in normal times.

CHAPTER 1 : HISTORY OF THE SURVEY.

Inception of the enquiry

1.1. Bengal was seriously affected by a famine in 1943, and various attempts were made to gauge the intensity of incidence, but mainly on general impressions. Material relating to the life of the rural people was meagre even for the period before the famine. Reliable data relating to the effect of the famine simply did not exist. A survey of a small number of villages was carried out by the Department of Anthropology of the University of Calcutta in November and December 1943 and a note embodying the results was issued by Professor K. P. Chattopadhyay, Head of the Department, early in February 1944. This survey was not based on the principle of random sampling. With the extremely limited resources at his disposal Prof. Chattopadhyay was naturally unable to do anything else than type studies on an extremely small scale ; no valid estimates for the province as a whole could possibly be based on the above survey.

1.2. In March 1944 K. P. Chattopadhyay approached P. C. Mahalanobis, Hony. Secretary, Indian Statistical Institute, with the suggestion that a survey on proper statistical lines should be organised as early as possible. It was decided that an extensive sample survey should be carried out by the Indian Statistical Institute under the general technical guidance of P. C. Mahalanobis with the collaboration of K. P. Chattopadhyay. The general plan of work was discussed in March and April 1944 ; and the training of investigators was started almost immediately. It was felt, however, that it would be advisable to secure necessary funds to enable the work being done on an adequate scale. On 12 May 1944 K. P. Chattopadhyay, in consulation with P. C. Mahalanobis, wrote to the Revenue Department, Government of Bengal for financial assistance, pointing out that the data collected by a survey of this type would enable a better planning of relief and rehabilitation, and eliminate wastage. A grant of Rs. 25000 was sanctioned by the Famine Relief Branch, Revenue Department, on 14 July 1944. The enquiry, which was being carried out at the expense of the Indian Statistical Institute, was expanded and the field survey on a provincial scale was started at the end of July 1944.

1.3. Field work was conducted under the direct guidance of K. P. Chattopadhyay. The general plan of the survey was prepared in the Indian Statistical Institute and the statistical analysis of the material was carried out nuder the guidance of P. C. Mahalanobis. Ramkrishna Mukherjea and Ambika Ghosh of the Indian Statistical Institute, who were in charge of the statistical work, maintained liaison between the field and statistical branches ; they also assisted in the training of field workers and the organisation of the field survey.

Progress of the survey

1.4. The original plan of field work was to complete the investigations during the three months of May, June and July 1944. The estimate of cost of the enquiry was prepared on this assumption. It had been contemplated that the whole of the field work would be done by voluntary workers taken from college students who would be able to do this work during the summer vacation which began in May and lasted till about the middle of July.

1.5. The delay in obtaining information about Government assistance, however, completely upset the above plan. By the time Government letter sanctioning assistance was received, it was already late in July. The services of student volunteers were no longer available. It became necessary therefore to obtain suitable volunteers from other sources. Contact was made with a number of organisations, and eventually the Bengal Provincial Kisan Sabha and certain local Congress organisations agreed to give the assistance of a number of workers ; 92 such volunteers were employed in different stages of the enquiry, the largest number having come from the B. P. K. Sabha. They were only given the expenses for their travel and board in addition to a small allowance to meet their out-of-pocket charges. We place on record our grateful appreciation for the services rendered by these volunteers and for the financial assistance given by the Government of Bengal.

1.6. The delay in receiving Government decision about assistance upset our original programme in other respects also. Rains had already set in when Government orders were received. Journeys which could have been previously undertaken on foot or by cycling. had now to be carried out by boats in many districts. The period allotted for journeys had thus to be lengthened to twice and often four times the original period. This seriously hampared and delayed the field work.

1.7. The main cultivation operations of the year had also started from July. This kept the agriculturists, who comprise the bulk of the rural population, extremely busy. Interviews had therefore to be spread over a longer period, delaying the progress still further. Finally, malaria set in from August and the field staff suffered acutely from this disease. No less than 80 of the investigators were attacked with malaria in the course of the survey, and about half of them had to drop out after doing a part of the work. Owing to these difficulties the field work had to be extended from the originally contemplated period of three months (May to July) to a little over six months from the end of July to the beginning of February, 1945. It was also not possible to take up the investigation of all the villages which it was originally intended to cover because it was neither feasible to obtain the services of voluntary workers for a much longer period nor to employ paid investigators for this purpose.

1.8. A third and still more important effect of the late decision of Government was that the very much increased cost of travel, and the increased cost involved in extending the investigation over six months resulted in the funds available for field work being exhausted before the total sample according to design could be surveyed. The Institute which, as would appear from subsequent paragraphs, had to meet, at least for the time being, a large part of the expenditure on statistical work from its own resources, was not in a position to finance the additional field work without further assistance from Government.

1.9. The analysis of the material was carried out in the Indian Statistical Institute. Hollerith sorting and tabulating units belonging to the Institute were used for this purpose which not only expedited the complicated analysis but diminished computational errors inherent in hand tabulation.

DELAY CAUSED BY LACK OF FUNDS

1.10. As noted in para 1.7. the field enquiry was completed by the beginning of February 1945, but the publication of the report has been delayed by more than one year. This has been mainly due to lack of funds. The cost incurred for the field enquiry (including preliminary expenses for the preparation of the design and printing of standard forms) was about Rs. 19,000. The total grant received from the Bengal Government was Rs. 25,000 so that a balance of only Rs. 6,000 was available for the statistical analysis. The field work itself had been carried out at the extremely low cost of only a little more than a rupee per family schedule. This was possible because a great deal of the work had been done by voluntary workers.

1.11. Specimen reports for two subdivisions (Tamluk in Midnapore district and Tangail in Mymensingh district) were submitted to the Government of Bengal on 22 February and 7 March 1945 respectively (immediately after the completion of the field work) with a request for instructions regarding the form for future tabulations. Mr. J. N. Talukdar, I.C.S., (then Director of Relief and Rehabilitation) and Mr. Symonds discussed this preliminary report with P. C. Mahalanobis and K. P. Chattopadhyay on 18 April 1945. In accordance with the suggestions made in the course of this discussion a full note on destitution (comprising practically chapter 3 of the present report) was submitted to Government on 8 June 1945 with a statement that over Rs. 20,000 had been spent on the statistical work by that time. A request was also made for additional funds to complete the analysis of the whole material. By this time Mr. Talukdar, who was taking a great deal of interest in the present enquiry, had however been transferred, and for a long time there was no response from Government.

1.12. In view of the importance and urgency of the work it was decided to carry on the statistical analysis with the available resources of the Indian Statistical Institute. It is estimated that statistical work to the value of Rs. 30,000 has been incorporated in the present report. A great deal of material still remains to be tabulated. Properly analysed it can throw much light on some of the major problems of rural Bengal such as indebtedness in relation to land, ownership of land, owner and share cultivation, comparative economic position of different castes and communities, age and sex distribution of rural population, etc Material relating to deaths during the famine was also collected, but the Relief Department did not want this information.

1.13. Very recently we received from Mr. S. K. Dey, I.C.S., Director of Relief and Rehabilitation, a letter dated 25 March 1946 in which he pointed out that the order of 14 July 1944 sanctioning the grant had "embodied the request that the results of the enquiry be reported to Government not later than the 30th September 1944". Mr. Dey also pointed out, that as this had not been done the Relief Department did not feel that there were grounds even "for approaching the Finance Department for further provision for this particular statistical enquiry".

1.14. The present enquiry has covered nearly 16,000 families in 386 villages scattered over the whole province, and was carried out to a very large extent, by voluntary help. The cost in similar enquiries already undertaken or proceeding in Bengal on behalf of Government varies from Rs. 10 or Rs. 12 to Rs. 20 or Rs. 25 per family schedule. The total cost to Government, if the present enquiry had been carried out in the usual way, would have been any thing between two and four lakhs of rupees. The date 30th September 1944 had been mentioned by us on the assumption that full scale field work would be started in May and completed by July. The Relief Department, however, took more than eight weeks (from 12 May to 14 July 1944) to sanction a grant in aid, but apparently still expected the work to be completed according to the original schedule. As already explained, the delay in receiving Government orders upset the whole programme of work and made it impossible to finish the field survey before the setting in of the monsoon. The position had been explained to Government in November 1944 but our difficulty was apparently not fully realised. In any case, in spite of repeated requests for funds to enable the Institute to proceed with the work, the Relief Department did not inform us until after 41 weeks (from 8 June 1945 to 25 March 1946) that the present enquiry was not considered deserving of further financial support from the Government of Bengal.

1.15. This report will give some idea of the kind of information which can be extracted from the material which was collected in the present enquiry. As already noted, a good deal of material still remains to be analysed. We were hoping that necessary financial help would be received from the Government of Bengal to enable us to complete the work. In view of the decision of the Revenue Department the report is being published in an incomplete form. The present report will, however, enable any one interested in the matter to judge whether enquiries of the present type have any value or not.

Arrangement of the report

1.16. The whole report is divided into two parts : (1) Statistical, and (2) Rehabilitation. The first part prepared by P.C. Mahalanobis, Ramkrishna Mukherjea and Ambika Ghosh[1], deals with the statistical aspects of the subject and gives a purely factual and technical account of the results of the sample survey. In this part suggestions from K. P. Chattapadhyay regarding tabutation and preparation of the report have been freely incorporated. The second part deals with various socio-economic problems created by the famine and necessary measures of rehabilitation, and has been prepared by K. P. Chattopadhyay and Ramkrishna Mukherjea[2]. Plans for rehabilitation given in the second part are based on the figures and results given in the statistical part. It must, however be clearly stated that the responsibility for the conclusions and views expressed in Parts 1 and 2 rests entirely on their respective authors.

---

[1] The first part is printed in this Journal.

[2] The second part has not been printed in *Sankhya*. The two parts are being published in the form of a book on *Famine and Rehabilitation in Bengal.*

CHAPTER 2 : DESIGN AND PROGRESS OF THE SURVEY.

DESIGN OF THE SURVEY

2.1. Famine conditions in 1943 had varied widely in different areas in Bengal. An all-Bengal view could have been obtained either (1) by surveying all the subdivisions[1], or(2) by stratifying Bengal into a few classes according to the degree of incidence of famine conditions and studying a number of subdivisions in each class. Obviously the first method would have given the most precise estimates for the province as a whole. But, owing to limited resources in money and men, this could not be attempted ; also the time taken to complete the enquiry would have been much longer. It was therefore decided to adopt stratification by intensity of incidence as sth iwould make it possible to assess the extent of disintegration in the distressed areas which was the main purpose of the enquiry.

2.2. The subdivisions in rural Bengal (86 in number) were classified by the Revenue Department according to the estimated intensity of incidence[2]; 12 were placed in Class I (most severely affected), 13 in Class II(*a*) (moderately affected), 12 in Class II (*b*) (less severely affected) and 6 in Class III (slightly affected); the remaining 43 subdivisions were stated to have been very slightly affected or not affected at all. It was decided to include in the survey all the subdivisions in Classes I, II and III. A few of the remaining 43 subdivisions (which may be referred to as Class IV) were to be surveyed as control to test the reliability of the classification. In each selected subdivision a number of villages were selected at random. These were of the order of 15 from each subdivision in Class I (roughly 1.5% of the total number of villages) ; about 10 from each subdivision of Class II (*a*) and Class II (*b*) (roughly 1% of the total number) ; about 8 from each subdivision of Class III (roughly 0.75% of the total number) ; and about 5 from each selected subdivision of Class IV (roughly 0.5% of the total).

2.3. Investigators were instructed to visit each of the selected village and first prepare a list of all resident families[3]. A sample of 20% of the families (but never less than 25 in absolute number) was next to be picked up at random in each village ; and each family included in the sample was to be then directly interrogated by the investigator and necessary entries made in standard schedules supplied to them. It may be mentioned that family composition was ascertained by the genealogical method.

SAMPLE SURVEYED

2.4. The original design covered 478 villages in 43 subdivisions excluding class IV ; and it had been intended also to include a few subdivisions from Class IV so as to include about 600 villages in all. But, as already noted in paragraph 1.7, it was not possible to complete the work for lack of funds. The result was that the survey had to be closed down after having covered 15,769 families from 386 villages in 41 subdivisions including Class IV. Excluding Class IV, the total number of subdivisions and villages surveyed were 38 and 371 respectively, that is, 88% and 78% of the number contemplated in the original design.

---

[1] An administrative unit covering an average area of about 850 square miles.

[2] *A Scheme for Relief and Rehabilitation in Bengal*, Revenue Department, Government of Bengal, 1944.

[3] In the present survey 'family' has been defined as the economic unit taking food from the same kitchen in conformity with the definition of 'occupied house' in the Census of 1931 and 1941.

2.5. The following list gives the 86 subdivisions of rural Bengal arranged according to the degree of incidence of famine conditions as classified by the Revenue Department. The number of villages included in the sample is shown within brackets in each case ; naturally no such number is given in the case of subdivisions excluded from the survey.

TABLE 2.1. SUBDIVISIONS CLASSIFIED ACCORDING TO DEGREE OF INCIDENCE OF FAMINE CONDITIONS AS CLASSIFIED IN *A Scheme for Relief and Rehabilitation in Bengal* (REVENUE DEPARTMENT, GOVERNMENT OF BENGAL 1944).

The number within brackets shows the number of villages included in the sample in each case ; no such number is given in the case of subdivisions excluded from the survey.

Class I—Most severely affected. 1. Arambagh (15), 2. Chandpur (6), 3. Chittagong (s)[1] (14), 4. Contai (15), 5. Daimond Harbour (15), 6. Madaripur (14), 7. Manikganj (7), 8. Munshiganj (14), 9. Nilphamari (14), 10. Noakhali (s) (10), 11. Tamluk (15), 12. 24-Parganas (s) (9).

Class II(a)—Moderately affected. 1. Barisal (s), 2. Bhola (10), 3. Coxbazar, 4. Dacca (s) (6), 5. Feni (8), 6. Goalanda, 7. Kandi (3), 8. Midnapur (s) (10), 8. Narayanganj (10), 10. Satkhira (10), 11. Serajganj (12), 12. Tangail (10) 13. Tipperah (s) (5).

Class II(b)—Less severely affected. 1. Bankura (s) (10), 2. Bishnupur (10), 3. Burdwan (s) (10), 4. Faridpur (s) (9), 5. Gaibandha (10), 6. Ghatal (10), 7. Hooghly (s) (10), 8. Howrah (s) (10), 9. Kalna (5), 10. Kusthia (8), 11. Pabna (s) (9), 12. Uluberia (10).

Class III—Slightly affected. 1. Bogra (s) (4), 2. Brahmanbaria, 3. Gopalganj, 4. Kishoreganj (8), 5. Kurigram (8), 6. Narail (8).

Class IV—Very slightly or not affected. 1. Alipur, 2. Asansol, 3. Bagherhat, 4. Balurghat, 5. Bangram, 6. Barasat (5), 7. Barrackpore, 8. Basirhat, 9. Birbhum (s), 10. Chittagong Hill Tracts, 11. Chuadanga, 12. Darjeeling (s), 13. Dinajpur (s), 14. Jalpaiguri (s), 15. Jamalpur, 16. Jangipur, 17. Jessore (s), 18. Jhargram. 19. Jhinaidah, 20. Kalimpong, 21. Khulna(s) 22. Katwa, 23. Kurseong, 24. Lalbagh, 25. Magura, 26. Maldah (s), 27. Meherpur (5), 28. Murshidabad (s), 29. Mymensingh(s), 30. Nadia (s)(5), 31. Naogaon, 32. Natore, 33. Netrokona, 34. Patuakhali, 35. Perojpur, 36. Rajshahi (s), 37. Ramgarh, 38. Rampurhat, 39. Ranaghat, 40. Rangpur (s), 41. Siliguri, 42. Serampore, 43. Thakurgaon.

2.6. The distribution of villages and subdivisions by the Revenue Department classification as included in the original design and as actually surveyed is shown below in Table 2.2

TABLE 2.2. NUMBER OF SUBDIVISIONS AND VILLAGES SURVEYED (BY REVENUE DEPARTMENT CLASSIFICATION)

Revenue Department classification	number of subdivisions		number of villages	
	in design	surveyed	in design	surveyed
(1)	(2)	(3)	(4)	(5)
Class I	12	12	180	148
Class II(a)	13	10	130	84
Class II(b)	12	12	120	111
Class III	6	4	48	28
Class IV	43	3	215	15
total	86	41	693	386

(s) indicates the Sadar subdivision of the district.

Quality of the material

2.7. In any socio-economic enquiry the quality of the primary material chiefly depends on the honesty of the investigator, and on his ability to ascertain the requisite information which in its turn depends on how far he succeeds in establishing contact and in overcoming the psychological resistance of the persons interviewed. It has been already noted that in the present enquiry most of the investigators had taken up the work from a sense of social duty; their honesty of purpose could, therefore, be thoroughly depended upon. In India the general public as a rule is not over anxious to confide to Government officials. Many of the volunteer-investigators had worked in relief and other organisations of social welfare; and sympathetic contact had been established with villagers in many districts in connexion with relief work organised by the People's Relief Committee of which K. P. Chattopadhyay is the Treasurer. The investigators were therefore generally successful in winning the confidence of the village people. The fact that they were working for a non-official organisation (and not a Government Department) was a distinct advantage. In some cases this might have, however, led the interviewees to exaggerate their sufferings. During the training period all the investigators had been cautioned against this bias, and the importance of recording correct facts for proper planning of rehabilitation was impressed on their minds. The use of a standard family schedule reduced the possibility of personal errors.

2.8. The primary schedules collected by the investigators were carefully scrutinized before being put up for analysis. Biswapada Das Gupta, M.Sc., and later Gautam Shankar Roy, M.Sc., both of whom had done post-graduate work in anthropology and had experience of field work, were in charge of the checking of the primary data under the direct supervision of K. P. Chattopadhyay. Some idea about the general quality of the material can be gathered from the fact that, as a result of such scrutiny, less than one percent of completed forms had to be rejected.

Nature of material included in the analysis

2.9. During analysis of any particular element (such as destitution, land transfer, etc.), family schedules which were suspected to be faulty, in respect of the element under analysis, were rejected from that particular analysis. This was done in order to obtain results based on reliable material for each item. For this reason, the total number of families studied for different elements is not always the same, and shows some minor variations. For destitution the total number of families included in the analysis is 15, 653; for land transfer it is 15,640; for cattle loss 15,242; and for change of occupation 15,208. The total variation in the size of the sample is thus not more than 2.8%, and is not likely to affect seriously the comparability of estimates in respect of different elements.

Occupational grouping of the data

2.10. Effect of the famine cannot be fully appreciated unless it is shown separately for different socio-economic sectors of the population. Family occupation was therefore adopted as the basis for classification. A family may, however, depend on two or more occupations of one or more earners of the family. In the present report, family occupation has been defined as that particular occupation from which the greater part of the total family income is derived.

2.11. Agricultural occupations have been sub-divided into four classes. The first group 'agriculture' represents mainly *ryots* (peasant-proprietor) who actually cultivate their own land or land taken on a share basis (*barga*) but who do not work as hired labourers. The second group 'agriculture & labour' consists of persons who not only cultivate their own and or land taken on a share basis but also work from time to time as agricultural labourers. The third group 'agricultural labour' consists of labourers who do not own any land or only a negligible area. The fourth group 'non-cultivating owner' is rather a heterogeneous one as it consists of *jotdars* who own large areas as well as widows and invalids who may belong to a different economic class but who are obliged to cultivate their land entirely by hired labourers or on a share basis.

2.12. The other occupations need no special explanation as the usual Census classification has been adopted. It should be noted, however, that the occupations under 'public administration and facilities and liberal arts' of the Census classification have all along been referred to in the report as 'profession & service', which in rural areas mainly consist of teachers in primary schools, unqualified doctors, priests, maulavis, petty clerks of the Union Board or village chowkidars, etc.

MULTIPLIERS FOR PROVINCIAL ESTIMATES

2.13. For lack of time it was not possible to adopt the area (or grid) method of sample survey in which an unequivocal multiplier can be used to obtain provincial estimates. Relevant Census figures for the total number of families (or persons) or the total number of plough cattle in rural Bengal (i.e. omitting urban areas) were therefore used to obtain provincial estimates. For such purposes, appropriate figures relating to the famine period were obtained by simple linear extrapolation from 1931 and 1941 Census data in regard to families or persons and 1930 and 1940 Census data in regard to plough cattle. In one or two cases the sample distributions were themselves also used for purposes of estimation.

UNRELIABILITY OF THE STRATIFICATION

2.14. It has been noted in praragraph 1.6 that the number of subdivisions and villages surveyed were 88% and 78% of the number in the original design. Although the entire sample according to design could not be covered, the total number surveyed is quite large and should have given fairly reliable results provided the stratification had been correct. Unfortunately, a first analysis of the material showed the Revenue Department grouping to be unreliable in as much as the rate of incidence of famine conditions in the different classes of subdivisions, as found by the sample survey, did not tally with the official classification. This can be seen from the following table 2.3 which gives the rates of destitution, land transfer, and cattle loss by classes of incidence according to the Revenue Department classification.[5]

TABLE 2.3. RATE OF DESTITUTION, LAND TRANSFER, AND CATTLE LOSS (ACCORDING TO REVENUE DEPARTMENT CLASSIFICATION).

Revenue Department classification of subdivision	percentage of total families: destitute in 1944	percentage of total families: selling cultivated paddy land in 1943	percentage of total families: losing plough cattle in 1943
(1)	(2)	(3)	(4)
Class I	4·5	10·6	2·6
Class II(a)	3·4	10·0	4·2
Class II(b)	2·5	7·6	3·0
Class III	3·1	11·9	3·7
Class IV	4·8	6·2	2·8

---

[5] 'Economic deterioration' was omitted from Table 2.3 as the estimates were based to some extent on a subjective interpretation of occupational data as noted in appendix 4(A-1).

The table indicates that the high rate of incidence in Class I regarding destitution and land transfer is in accordance with the classification of the Revenue Department ; the proportion of families losing plough cattle is, however, lower than that in all other classes. In respect of all three elements, rates in Class III are higher than those in Class II(b). A still more surprising anomaly is seen in Class IV which shows highest rates of incidence in destitution, and quite high rates in the other two elements although the subdivisions in this class were considered by the Revenue Department as having been unaffected by famine conditions.

2.15. It is true that some of the worst affected families had migrated to cities so that a village might have been severely affected ; and yet, owing to the withdrawal of some of those very families which had been most hard hit by famine conditions, a survey conducted in 1944 might not have necessarily shown a servere degree of incidence. This may to some extent explain the discrepancy between the Revenue Department classification and the results of the sample survey. It must be stated however that the field staff had been instructed to note carefully the fact of any family which had been included in the sample but which had become extinct ; the number of such cases were small. Not only this ; subdivisions which were considered as not affected in the Revenue Department classifications were actually found to be appreciably or even severely affected by the sample survey. Withdrawal of extinct families can not possibly explain this type of discrepancy. In any case, it is clear that during the period covered by the sample survey the Revenue Department classification was not in accord with the real situation in the country[6]. This was a serious setback. For purposes of estimation, other classifications had therefore to be tried ; and finally that given by the Director of Industries was adopted[7]. This was found to be most in accord with the results of the sample survey (as will be seen from the later chapters), and all estimates in the present report are therefore based on this classification.

2.16. The following table 2.4 gives the 86 subdivisions of rural Bengal arranged according to the degree of incidence of famine conditions as classified by the Director of Industries, Government of Bengal. As in table 2.1, the number of villages included in the sample is shown within brackets in each case ; no such number is given in the case of subdivisions excluded from the survey.

TABLE 2.4. SUBDIVISIONS ARRANGED ACCORDING TO DEGREE OF INCIDENCE OF FAMINE CONDITIONS IN *A plea for the Rehabilitation of Bengal's Rural & Industrial Economy* (DEPARTMENT OF INDUSTRIES).

Class A (*Very severely affected*) 1. Brahmanbaria, 2. Chandpur (6), 3. Comilla (5), 4. Contai (15), 5. Dacca (s) (6), 6. Diamond Harbour (15), 7. Faridpur(s) (9), 8. Feni (8), 9. Goalanda, 10. Madaripur (14), 11. Munshiganj (7), 12. Manikganj (10), 13. Narayanganj (10), 14. Noakhali (s) (10), 15. Tamluk (15).

Class B (*Severely affected*) 1. Bagerhat, 2. Bakharganj(s), 3. Balurghat, 4. Bhola (10), 5. Bongaon, 6. Chittagong (s) (14), 7. Jamalpur, 8. Jessore (s), 9. Jhinaidha, 10. Kandi (3), 11. Khulna (s), 12. Kishoreganj (8), 13. Magura, 14. Mymensingh (S), 15. Narail (8), 16. Natore, 17. Netrokona, 18. Nilphamari (14), 19. Pabna (s) (9), 20. Patuakhali, 21. Pirojpur, 22. Satkhira (10), 23. Serajganj (12), 24. Tangail (10), 25. Uluberia (10).

Class C-1 (*Partially affected*) 1. Alipore, 2. Darjeeling (s), 3. Jalpaiguri (s), 4. Kalimpong, 5. Kurseong, 6. Siliguri.

Class C-2 (*Slightly affected*) 1. Arambagh (15), 2. Asansol, 3. Bankura (s) (10), 4. Barasat(5), 5. Barrakpore, 6. Basirhat, 7. Bishnupur (10), 8. Bogra(s) (4), 9. Burdwan (s) (10), 10. Chuadanga, 11. Cox Bazar, 12. Dinajpur (s), 13. Gaibandha(10), 14. Ghatal(10), 15. Gopalganj, 16. Hooghly(s)(10), 17. Howrah(s)(10), 18. Jangipur, 19. Jhargram, 20. Kalna (5), 21. Katwa, 22. Kurigram (8), 23. Kustia (8), 24. Lalbagh, 25. Maldah (s), 26. Meherpur (5), 27. Midnapore, (s) (10) 28. Murshidabad (s), 29. Nadia (s) (5), 30. Naogaon, 31. Rajshahi (s), 32. Rampurhat, 33. Ramgarh, 34. Ranaghat, 35. Rangamati, 36. Rangpur (s), 37. Serampore, 38. Suri, 39. Thakurgaon 40. 24 Parganas (s) (9).

---

[6] Incidentally this shows that reports collected through the district officers can be highly erroneous as these are often based on the subjective impressions of a few officials.

[7] *A Plea for the Rehabilitation of Bengal's Rural and Industrial Economy* 1944 (Department of Industries).

2.17. The distribution of villages and subdivisions by the Industries Department classification has been shown in table 2.5 separately for the number intended to have been covered in the original design and the number which was actually surveyed.

TABLE 2.5. NUMBER OF SUBDIVISIONS AND VILLAGES INCLUDED IN THE SURVEY (BY INDUSTRIES DEPARTMENT CLASSIFICATION).

Industries Department classification	subdivisions		villages	
	in design	surveyed	in design	surveyed
(1)	(2)	(3)	(4)	(5)
Class A	15	13	225	134
Class B	25	11	250	108
Class C	46	17	248	144
total	86	41	723	386

According to the present classification the number of subdivisions surveyed is 13 out of 15 in Class (A), 11 out of 25 in Class (B), and 17 out of 46 in Classes (C-1) and (C-2) which have been pooled together to give one single Class (C). The number of villages surveyed is 134 in Class (A), 108 in class (B), and 144 in class (C).

2.18. The accompanying sketch map of Bengal shows the subdivisions hatched according to degree of incidence of famine conditions (Industries Department classification). The intensity of sampling is shown by stars ; a single star (*) indicates that from 1 to 5 villages were included in the survey ; two stars (**) that from 6 to 10 and three stars (***) that from 11 to 15 villages were so included. A list of the names of subdivisions according to the serial numbers on the map is given in Table 2.6. The actual number of villages and the number of families covered in the survey are also shown against each subdivision. It will be noticed that the middle belt and eastern regions of the province were adequately covered in the survey ; considerable portions of North and West Bengal had been, however, excluded.

LIMITATIONS OF THE DATA

2.19. It is obvious from the design of the survey that estimates for Classes (A) and (B) would be fairly reliable as the number of subdivisions included in the sample from these two strata cover a good proportion of the subdivisions. For Class (C), the number of subdivisions covered is however comparatively small and the stratum was not adequately covered and the sub-divisions were also not selected strictly on a random basis. As already noticed, there are considerable gaps in the north and the west of the province. Estimates for Class (C) must be therefore necessarily less reliable. The all-Bengal estimates would also be affected because of this weakness in the design in regard to Class (C).

2.20. In this situation it has been considered advisable to give separate estimates for the 41 subdivisions included in the survey. The margin of error of such estimates have been calculated in the case of important items and are reasonably small. In any case, the problem of rehabilitation being essentially that of the affected areas, results based on the present survey are likely to give sufficiently reliable information for many practical purposes. Estimates for the province as a whole have been also given. In view of the fact that 41 out of 86 subdivisions or nearly half the province had been covered, these estimates probably give a fairly good dimensional idea of the actual situation.

Reliability of provincial estimates

2.21. Some idea of the accuracy of the provincial estimates made in this report can be obtained by comparing the results based on the present sample survey with estimates previously ascertained in other ways. As already explained, in the present report the provincial estimates are mostly based on the family as the unit. Thus table 2.7 below gives the estimated total number of plough cattle and the total acreage of paddy land in rural Bengal in April 1943 as calculated from the mean number of plough cattle and mean acreage of paddy land owned per family in the sample (with appropriate weightage in accordance with the stratification by the Industries Department classification).

TABLE 2.7. ESTIMATED TOTAL NUMBER (IN LAKHS[8]) OF PLOUGH CATTLE AND PADDY LAND (IN LAKHS OF ACRES) IN RURAL BENGAL IN APRIL 1943 OBTAINED FROM THE SURVEY WITH WEIGHTAGE BASED ON INDUSTRIES DEPARTMENT CLASSIFICATION.

Industries Department classification	total number of families in January 1943 (in lakhs)	mean value per family in April1943		estimates in lakhs in April 1943	
		plough cattle	paddy land in acres	plough cattle	paddy land in acres
(1)	(2)	(3)	(4)	(5)	(6)
Class A	25·9	0·58	1·74	15·02	45·07
Class B	33·9	0·78	1·91	26·44	64·75
Class C	42·6	0·88	1·80	37·49	76·68
Rural Bengal	102·4	0·77	1·82	78·95	186·50

From the bottom row of col. (5) it is seen that the total number of plough cattle in rural Bengal in April 1943 was about 79 lakhs. Extrapolating from the cattle Census figures of 1930 (99.89 lakhs) and 1940 (86.99 lakhs) the calculated number of plough cattle in rural Bengal in the beginning of 1943 is found to be about 83 lakhs (para 6.2 of the present report). The difference between the extrapolated value and the sample estmate is about 4 lakhs or about 5% of the estimate based on the cattle census.

2.22. Again from col. (6) the total area of paddy land in rural Bengal is found to be about 187 lakhs of acres in April 1943 according to the sample survey. The total area under *aman* paddy (the main winter crop) as given in the official estimate of the Agricultural Department of the Government of Bengal was 182 lakhs of acres in 1942-43. According to the crop survey carried out by the Indian Statistical Institute, the area sown with *aman* paddy was about 240 lakhs of acre (with a margin of error of 10%) in 1943-44 in the season immediately following the famine, about 222 lakhs of acre in 1944-45 one year later, and 212 lakhs of acre in 1945-46 ; the effective margin of error in both the latter cases being of the order of 3% or 4%. It is known that owing to the abnormal increase in the price of rice during the famine there was a large increase of acreage under paddy which was brought about by sowing paddy on marginal, fallow or other lands on which paddy is not normally cultivated.

2.23. Ambiguities in the meaning of the phrase 'amount of paddy land' make exact comparisons difficult. In Bengal the two important paddy crops are (winter) *aman* comprising about 77% and (monsoon) *aus* comprising nearly 23% on an average of the total crop (neglecting a very small quantity of *boro* or summer paddy). *Aus* and *aman* are, however, often grown on the same land. 'The amount of paddy land' cannot, therefore, mean the total area on which either or both *aman* or *aus* paddy is grown. Then there are large seasonal fluctuations in the amount of land on which paddy is actually grown depending to a large extent on the amount and distribution of rainfall, and partly also on prices of agricultural commodities and other economic factors. The 'amount of paddy land', therefore, probably

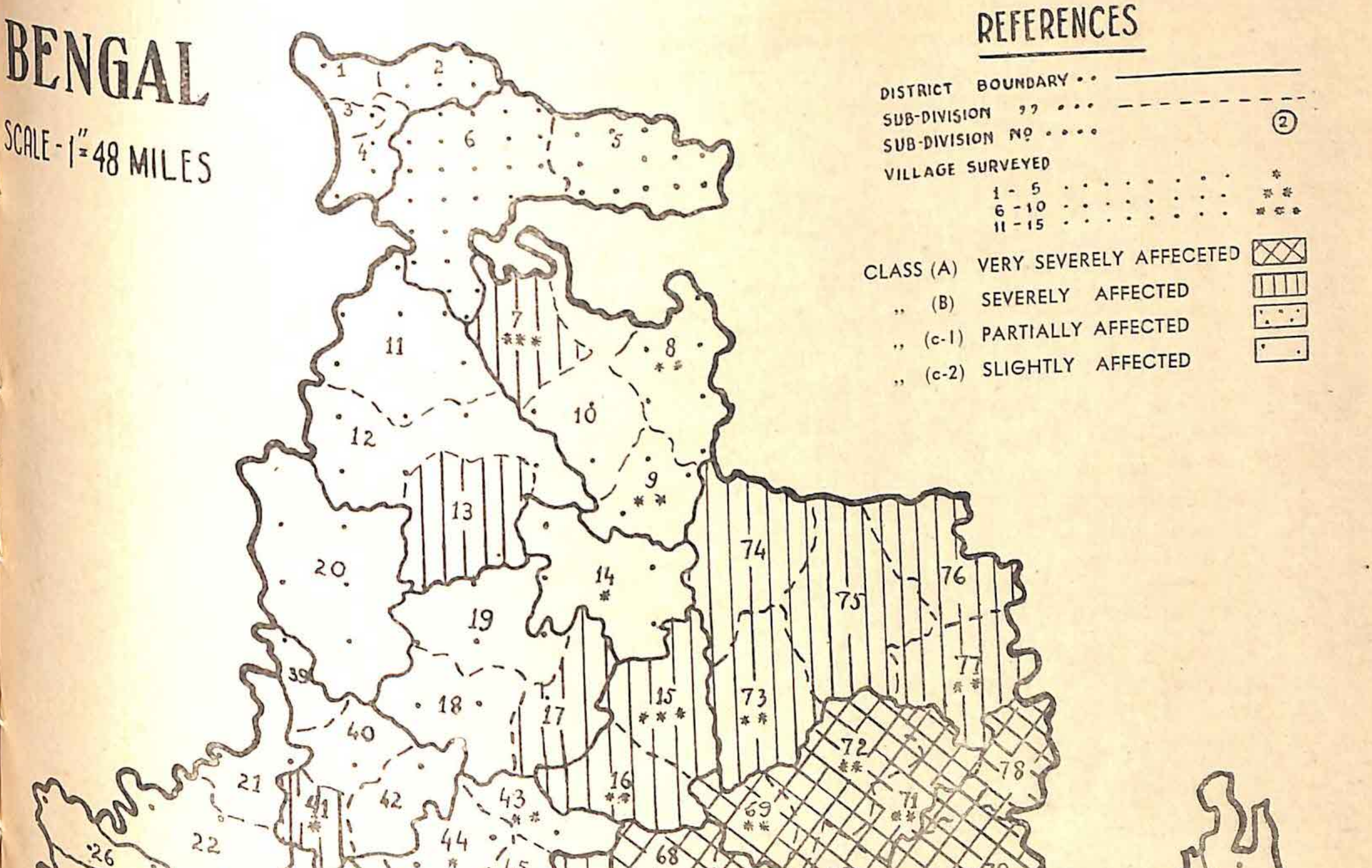
---

[8] 1 lakh=100,000=0·1 million.

Table 2·6 List of subdivisions according to serial numbers on the map with intensity of sampling and number of villages and families included in the survey in each subdivision.

subdivision	intensity of sampling	number surveyed		subdivision	intensity of sampling	number surveyed	
		village	family			village	famiy
1. Darjeeling (s)	—	—	—	44. Meherpur	*	5	165
2. Kalimpong	—	—	—	45. Chuadanga	—	—	—
3. Kurseong	—	—	—	46. Nadia (s)	*	5	154
4. Siliguri	—	—	—	47. Ranaghat	—	—	—
5. Alipur Duars	—	—	—	48. Barrackpore	—	—	—
6. Jalpaiguri (s)				49. Barasat	*	5	146
7. Nilphamari	***	14	980	50. Diamond-Harbour	***	15	543
8. Kurigram	**	8	281	51. 24-Parganas (s)	**	9	322
9. Gaibandha	**	10	450	52. Basirhat	—	—	—
10. Rangpur (s)	—	—	—	53. Bangram	—	—	—
11. Thakurgaon	—	—	—	54. Jhinaidah	—	—	—
12. Dinajpur (s)	—	—	—	55. Magura	—	—	—
13. Balurghat	—	—	—	56. Narail	**	8	229
14. Bogra (s)	*	4	121	57. Jessore (s)	—	—	—
15. Sirajganj	***	12	290	58. Satkhira	**	10	
16. Pabna (s)	**	9	238	59. Khulna (s)	—	—	—
17. Natore	—	—	—	60. Bagherhat	—	—	—
18. Rajshahi (s)	—	—	—	61. Perojpur	—	—	—
19. Naogaon	—	—	—	62. Patuakhali	—	—	—
20. Maldah (s)	—	—	—	63. Bhola	**	10	485
21. Rampurhat	—	—	—	64. Barisal (s)	—	—	—
22. Birbhum (s)	—	—	—	65. Madaripur	***	14	370
23. Katwa	—	—	—	66. Gopalganj	—	—	—
24. Kalna	*	5	120	67. Faridpur (s)	**	9	372
25. Burdwan (s)	**	10	350	68. Goalundo	—	—	—
26. Asansol	—	—	—	69. Manikaganj	**	10	187
27. Bishnupur	**	10	320	70. Munshiganj	**	7	633
28. Bankura (s)	**	10	264	71. Narayanganj	**	10	369
29. Jhargram	—	—	—	72. Dacca (s)	**	6	167
30. Midnapur (s)	**	10	254	73. Tangail	**	10	365
31. Contai	***	15	521	74. Jamalpur	—	—	—
32. Tamluk	***	15	599	75. Mymensingh (s)	—	—	—
33. Ghatal	**	10	353	75. Netrokona	—	—	—
34. Ulubaria	**	10	384	77. Kishorgunj	**	8	383
35. Howrah (s)	**	10	405	78. Brahmanbaria	—	—	—
36. Arambagh	***	15	435	79. Tippera (s)	*	5	122
37. Serampore	—	—	—	80. Chandpur	**	6	249
38. Hooghly (s)	**	10	242	81. Noakhali (s)	**	10	773
39. Jhangipur	—	—	—	82. Feni	**	8	348
40. Lalbagh	—	—	—	83. Chittagong (s)	***	14	1931
41. Kandi	*	3	149	84. Cox's Bazar	—	—	—
42. Murshidabad (s)	—	—	—	85. Ramgarh	—	—	—
43. Kustia	**	8	284	86. Rangamati	—	—	—

(s) indicates the Sadar subdivision of the district.



Sketch map of Bengal showing subdivisions hatched according to degree of incidence of famine conditions (Industries Department classification) with intensity of sampling in subdivisions shown by stars.

means the land on which paddy can be grown in normal years, and would not include marginal land on which paddy can be grown only in exceptional years of very good rainfall. The sample estimate of 186 lakhs of acre in the year previous to the famine is thus not inconsistent with other available estimates.

2.24. One further point deserves mention in the present connexion. From cols. (3) and (4) of table 2.7 it is seen that the average number of plough cattle or acres of paddy land owned per family were lower in Class (A)subdivisions which were more severely affected by the famine than in Class (B) or Class (C) subdivisions. This is, of course, just what is to be expected. Subdivisions in which rural families owned less plough cattle or less paddy land were obviously more vulnerable to famine conditions. It would be also remembered that subdivisions which were believed to have been less affected by the famine had been to a large extent left out of the sampling scheme. Regions which were omitted were thus less susceptible to the famine. It is likely, therefore, that these were the regions in which families had owned, on an average, a larger number of plough cattle or a larger amount of paddy land. Omission of such regions would lower the calculated averages of plough cattle and paddy land per family. On this view, the present provincial estimates of total plough cattle and total paddy land are likely to be underestimates.

2.25. Another comparison is possible, namely, between the age distribution of the population by sexes in the present sample and the corresponding distribution for the Census population of 1931 (this because age-tables for 1941 were not calculated and are not available). The age distribution of the population in the sample shows a definite and comparatively large excess of persons of both sexes below 20 years of age in comparison with census figures of 1931. This may have arisen to some extent by the emigration of adults from rural to urban areas caused partly by a phenomenal increase in the demand for both skilled and unskilled labour in various kinds of work connected with the war, and partly by the growing economic distress in rural areas during the famine period. It is, however, not possible to state whether the observed excess of young persons below 20 years of age can be adequately explained by the two factors mentioned above, or the sample figures are subject to unknown errors.

Margin of error

2.26. The margin of error of provincial or other estimates has been calculated from the sample material in a number of cases to which reference has been made at appropriate places. For convenience of reference, it may be stated here that for the rate of destitution in May 1944 (percentage of total families destitute) the estimated value is 3.35% with a standard error of 0.16% (Table 3.1) or a percentage margin of error[9] of 4.8% ; the estimated value of the percentage of families selling paddy land between April 1943 and April 1944 is 9.0% with a standard error of 0.27% (Table 5.4) or a margin of error of 3% ; and the estimated value of the percentage of families mortgaging paddy land between April 1943 and April 1944 is 6.3% with a standard error of 0.24% (Table 5.6) or with a margin of error of 3.8%. It will be noticed that the margin of error of provincial estimates varies from 3% to 4.8% and is quite small. Estimates for smaller units (like the district or the sub-division) would be, of course, higher. But, on the whole, the results of the sample survey (especially estimates for the 41 subdivision covered in the survey) supply a farily reliable picture of economic conditions after the famine of 1943 ; this, of course, within the limits of sampling fluctuation in respect of all the results discussed in the present report.

---

(9) Standard error of the mean is expressed as a percentage of the mean valuei tself.

CHAPTER 3. DESTITUTION

3.1. The famine of 1943 left a large number of destitutes in its trail. For a quantitative study a destitute family has been defined in this report as one which is either solely or primarily (in cases where there used to be some ostensible means of occupation before the famine but which was no longer available at the time of the survey) dependent on charity.

3.2. *Proportion of destitute families before and after famine.* Table 3.1 shows the rate of destitution by families before and after the famine.

TABLE 3.1. NUMBER AND PROPORTION OF FAMILIES LIVING ON CHARITY IN JANUARY 1943 AND IN MAY 1944.

		destitute families living on charity									
		number surveyed					percentage of total number of families				
class	total number of families surveyed	total in January 1943	change between January 1943 and May 1944		total in May 1944	net increase between Jan.1943, and May 1944	total in January 1943	change between January 1943 and May 1944		total in May 1944	net increase between Jan.1943 and May 1944
			increase	decrease				increase	decrease		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Class A	5089	150	126	26	250	100	2·95	2·45	0·51	4·92	1·97
Class B	5769	141	80	36	185	44	2·44	1·39	0·62	3·21	0·77
Class C	4795	91	38	4	125	34	1·90	0·79	0·08	2·61	0·71
rural Bengal	15653	382	244	66	560	178	2·34	1·42	0·37	3·39	1·05
41 subdivisions surveyed										3·35±0·16	

Weights used for the provincial estimates are the total number of families in January 1943 in each class as found by extrapolation of the figures of Census of 1931 and 1941.

Cols. (8) and (11) of the table indicate that the estimated provincial rates of destitution by families before and after the famine were 2.34% and 3.39% respectively. Col.(11) shows that the estimated rate of destitution in May 1944 for the 41 subdivisions surveyed was 3.35% with a standard error of ±0.16 only. (This of course, leaves out of consideration families which had become extinct and were not covered by the survey ; but, as already noted, the total number of such families could not have been very large).

3.3 It will be noticed from col.(12) of the table that there was an appreciable increase in the proportion of destitute families in all classes of subdivisions between January 1943 and May 1944. Col. (11) shows that the proportion of destitute families in May 1944 was 4.92% or highest in Class(A) ; 3.21% in Class(B) ; and 2.61% in Class(C) which agree generally with the scheme of classification. It will be also noticed from col. (8) that proportions of destitute families were originally arranged in exactly the same order in the three different classes in 1943. The parallel grading in January 1943 and May 1944 probably indicates that by the beginning of 1943 adverse conditions created by war had already affected the different areas broadly in the same pattern as in 1944.

3.4. *Proportion of destitute persons before and after famine.* Table 3.2 shows the rate of destitution by persons before and after the famine.

Table 3.2. Number and proportion of destitute persons living on charity in January 1943 and in May 1944 and change between January 1943 and May 1944.

class	total number of persons surveyed	destitute persons living on charity									
		number surveyed					percentage of total number of persons				
		total in January 1943	change between January 1943 and May 1944		total in May 1944	net increase between Jan.1943 and May 1944	total in January 1943	change between January 1943 and May 1944		total in May 1944	net increase between Jan. 1943 and May 1944
			increase	decrease				increase	decrease		
(1)	(2)	(3)	(4)	(5	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Class A	27073	454	478	12	820	366	1·68	1·76	0·41	3·03	1·35
Claes B	31452	427	216	146	497	70	1·36	0·68	0·46	1·58	0·22
Class C	21954	240	118	17	341	101	1·09	0·52	0·08	1·53	0·44
rural Bengal	80479	1121	812	275	1658	537	1·35	0·90	0·30	1·95	0·60

Weights used for the provincial estimates are the total population in January 1943 in each class as found by extrapolation of the figures of Census of 1931 and 1941.

Cols.(8) and (11) of the table indicate that estimated provincial rates of destitution by persons before and after the famine were 1.35% and 1.95% respectively. Col. (12) of the table shows that, as in the case of destitute families in table 3.1 there was an appreciable increase in the proportion of destitute persons in all classes of subdivisions between January 1943 and May 1944. Col. (11) of table 3.2 shows similar features to the corresponding column of table 3.1 indicating broad agreement in the classification.

3.5. *Adverse conditions in Bengal since the end of* 1942. It will be noticed from col. (8) of tables 3.1 and 3.2 that the proportions of destitute families and persons in January 1943 were in broad agreement with the present classification. Adverse conditions had apparently already set in by the end of 1942. This is corroborated by the fact that the proportion of destitute persons in January 1943, as estimated by the present sample survey, was 1.35% against a general average of 1.07% of persons living on charity in 1931 as given in the Census Report (the figure for 1941 is not available).

3.6. *Total number of destitutes before and after famine.* Table 3.3 gives the estimated number of destitute families and persons in January 1943 and May 1944.

Table 3.3. Estimated number (in lakhs) of destitute families and persons in rural Bengal in January 1943 and May 1944.

class	estimated number in lakhs of destitutes					
	families			persons		
	January 1942	May 1944	change between Jan. 1943 & May 1944	January 1943	May 1944	change between Jan. 1943 & May 1944
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Class A	0·76	1·27	0·51	2·50	4·51	2·01
Class B	0·83	1·09	0·26	2·68	3·07	0·39
Class C	0·81	1·11	0·30	2·27	3·18	0·91
rural Bengal	2·40	3·47	1·07	7·45	10·76	3·31

It will be noticed from cols. (2) and (5) of table 3.3 that at the beginning of 1943 roughly about 2.4 lakhs of families comprising about 7.5 lakhs of persons were destitutes. Cols. (3) and (6) show that these numbers had increased to roughly 3.5 lakhs of destitute families and 10.8 lakhs of destitute persons in May 1944. That is, about 1.1 lakhs of families comprising 3.3 lakhs of persons had become destitute owing to the famine of 1943.

3.7. *Destitution due to war and famine conditions in Bengal*. If we assume that the Census proportion of 1.07% of destitute persons in 1931 had remained steady, and might therefore be accepted as a 'normal rate' then the number of destitute persons (on the basis of a rural population of 55.2 millions[1] at the beginning of 1943) should have been about 6.0 lakhs in January 1943. The estimated number was about 7.5 lakhs; it seems therefore, that at the beginning of 1943 there had been already an increase of 1.5 lakhs of destitute persons. There was a further increase of about 3.3 lakhs of destitute persons between January 1943 and May 1944. At the normal rate of destitution of 1.07%, the normal number in May 1944 should have been about 6.0 lakhs; the estimated figure was about 10.8 lakhs showing that probably 4.8 lakhs of people had become destitutes under war and famine conditions in Bengal.

3.8. *Age and sex of the destitutes before and after famine*. Table 3.4 gives the estimates of destitute persons in January 1943 and May 1944 and the new destitutes created by the famine in the intervening period by age and sex.

TABLE 3.4. ESTIMATED NUMBER (IN LAKHS) OF DESTITUTE PERSONS IN RURAL BENGAL BY AGE AND SEX IN JANUARY 1943 AND IN MAY 1944 AND THE NUMBER OF NEW DESTITUTES BETWEEN JANUARY 1943 AND MAY 1944.

age group (in years)	estimated number of destitute persons						estimated number of new destitute persons (in lakhs) between Jan. 1943 & May 1944		
	January 1943			May 1944					
	male	female	total	male	female	total	male	female	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0— 5	0·27	0·31	0·58	0·60	0·50	1·10	0·33	0·19	0·52
5—15	1·61	1·06	2·67	2·19	1·68	3·87	0·58	0·62	1·20
15—50	0·98	1·95	2·93	1·43	2·92	4·35	0·45	0·97	1·42
above 50	0·50	0·77	1·27	0·59	0·85	1·44	0·09	0·08	0·17
total	3·36	4·09	7·45	4·81	5·95	10·76	1·45	1·86	3·31

This is a summary table from table 3A-1 (given in the appendix) which gives the distribution by classes of incidence.

It will be noticed from col. (10) of table 3.4 that the total increase in the number of destitutes of age (0—5) years was 52,000; between (5—15) years, 1,20,000; between (15—50) years 1,42,000; and of age 50 years and above, 17,000. Proportionately to actual numbers the increase was thus heaviest in the younger age-groups.

3.9. Cols. (8) and (9) of the table show that in the age-group (15—50) the number of new destitute women is double the number of new destitute men, while in the age-groups (5—15) and (above 50) both are nearly the same and in the age-group (0—5) the number of destitute men is nearly double that of women. The over all figures indicate that the number of destitute women is greater than that of destitute men. The large number of destitute women of adult age (15—50) has created a serious socio-economic problem for the country.

---

[1] This is the extrapolated figure obtained from the 1931 and 1941 Census population of 45.46 and 53.44 millions respectively in rural areas of Bengal. (1 million = 10 lakhs).

3.10. *Classification of destitutes by occupational groups before famine.* The number of destitute families and persons in the subdivisions surveyed living on charity in May 1944 and classified by family occupation in January 1943 is given in table 3.5. Because of the manner in which this table has been formed, the number of families and persons given in each occupational class (except, of course, in the group living on charity) actually show the number of new destitutes, that is, the number of families or persons who became destitute between January 1943 and May 1944.

TABLE 3.5. ESTIMATED TOTAL NUMBER (IN LAKHS) OF DESTITUTE FAMILIES AND PERSONS BY OCCUPATIONAL GROUPS AS IN JANUARY 1943.

occupational groups as in January 1943	destitutes living on charity in May, 1944					
	number surveyed		percentage of total		estimated number (in lakhs)	
	families	persons	families	persons	families	persons
(1)	(2)	(3)	(4)	(5)	(6)	(7)
agriculture	41	126	7·3	8·4	0·25	0·90
agriculture & labour	23	78	4·1	4·5	0·14	0·49
agricultural labour	82	256	14·7	18·5	0·51	1·99
non-cultivating owner	10	34	1·8	2·5	0·06	0·27
husking paddy	8	13	1·4	0·6	0·05	0·06
fishing	17	71	3·1	4·2	0·11	0·45
craft	22	78	3·9	5·2	0·14	0·56
transport	3	11	0·5	0·7	0·02	0·08
trade	14	46	2·5	3·7	0·09	0·40
profession & service	15	72	2·7	4·1	0·09	0·44
non-agricultural labour	3	11	0·5	1·0	0·02	0·11
other productive occupations	6	14	1·1	1·3	0·04	0·14
living on charity in January 1943	316	848	56·4	45·3	1·95	4·87
total	560	1658	100·0	100·0	3·47	10·76

It will be noticed from col. (7) of the table that 'agricultural labour' contributed the largest share (about 2 lakhs) of destitutes ; 'agriculture' contributed 90,000 ; and 'agriculture & labour' 49,000. The total share of all agricultural occupations was thus very heavy, 3,65,000 of destitutes. 'Crafts' contributed 56,000; while 'fishing' and 'profession & service' each contributed about 45,000 ; 'trade' also had quite a good share, namely, 40,000 ; 'husking paddy', 'transport', and 'non-agricultural labour' on the other hand each contributed only about 11,000 or less.

3.11. *Comparative incidence in the occupational groups.* The above figures, however, do not show comparative rates of incidence in different occupational groups. Thus the large number of destitutes from agricultural groups compared to the meagre number from 'fishing' apparently suggests that the latter group was less affected. But this might not have been so in fact. The total population in agricultural groups is large while the total number having fishing as their occupation is very small, so that the actual number of destitutes in agricultural groups may be easily large even when the severity of famine conditions is much greater in 'fishing' or such other groups. This can be easily seen from table 3.6 which gives the proportional distribution by occupations of rural families living in January 1943 based on the present sample and corresponding proportions of families turned destitute between January 1943 and May 1944 in each occupational group.

TABLE 3.6. RELATIVE PROPORTION OF FAMILIES DESTITUTE IN MAY 1944 TO THE PROPORTION OF FAMILIES WITHIN EACH OCCUPATIONAL GROUP AS IN JANUARY 1943.

	proportion of families		
occupational groups as in January 1943	within each occupational group in January 1943	destitute in May 1944	index of destitution $\left(\frac{100 \times \text{col. (3)}}{\text{col. (2)}}\right)$
(1)	(2)	(3)	(4)
agriculture	32·67	7·3	22
agriculture & labour	16·75	4·1	24
agricultural labour	16·52	14·7	89
non-cultivating owner	5·99	1·8	30
husking paddy	1·78	1·4	79
fishing	1·39	3·1	223
craft	5·06	3·9	77
transport	0·67	0·5	75
trade	6·71	2·5	37
profession & service	6·77	2·7	40
non-agricultural labour	0·98	0·5	51
other productive occupations	2·13	1·1	52
living on charity in January 1943	2·58	56·4	—
total	100·00	100·0	—

It is seen from col. (2) of this table that 32.67% of all families included in the survey had 'agriculture' as their occupation in January 1943, while the percentage share of destitutes in May 1944 contributed by the 'agriculture' group was 7.3% as shown in col. (3). The relative proportion or index of destitution, obtained by expressing 7.3% as a percentage of 32.67%, is thus 22 as shown in col.(4). In the same way, 'fishing' had contributed only 1.39% of total families but no less than 3.1% of destitute families ; the relative contribution of index number of destitution is thus 223 which is very high. Col. (4) of the table shows that the occupational groups arranged in a decreasing order of index of destitution, i.e. of serverity of incidence were, (1) fishing (223), (2) agricultural labour (89), (3) husking paddy (79), (4) craft (77), (5) transport (75) and then the remaining occupations.

3.12. *Cause of destitution before and after famine.* Table 3.7 gives the distribution of destitute persons by the condition of earners of destitute families (classified as dead, emigrated, invalid or unemployed) and thus indicates the immediate cause of destitution.

Table 3.7. Persons living on charity in May 1944 with productive or unproductive occupations in January 1943 by cause of destitution, i.e. condition of parents & earners

condition of earners	persons living on charity in May 1944 with occupations in January 1943							
	number surveyed			percentage of total number of destitute		estimated total number (in lakhs)		
	living on charity	with productive occupation	total	living on charity	with productive occupation	living on charity	with productive occupation	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
earners dead	415	394	809	25·03	23·77	2·69	2·51	5·20
earners emigrated	4	5	9	0·24	0·30	0·03	0·03	0·06
sub-total (1)	419	399	818	25·27	24·07	2·72	2·54	5·26
earners invalid	263	203	466	15·86	12·24	1·71	1·37	3·08
earners unemployed	164	210	374	9·9	12·67	1·06	1·36	2·42
sub-total (2)	427	413	840	25·75	24·91	2·77	2·73	5·50
total	846	812	1658	51·02	48·98	3·49	5·27	10·76

Cols.(5) and (6) of the table show that there is no large difference between the proportion of destitutes under different condition of earners in pre and post-famine periods. Col.(9) shows that death of earners accounts for almost half the cases of destitution ; next in importance is sickness and then unemployment of earners.

3.13. *Summary of conclusions.*

(1) Adverse conditions leading to an increase in the number of destitutes appear to have set in before 1943.

(2) About 4.8 lakhs of people had become destitutes under war and famine conditions in Bengal ; and of this number about 3.3 lakhs had probably been contributed by the famine of 1943.

3 The total number of post-famine destitutes was of the order of 10.8 lakhs in May 1944.

4 Compared to the number of destitutes in January 1943 the largest proportion had come from the younger age-groups during the famine period of January 1943 and May 1944.

(5) The proportion of destitute women is greater than the destitute men and a large number of them are adults belonging to the age group of 15—50 years. This has created a serious socio-economic problem for the country.

(6) In actual numbers, the occupation of 'agricultural labour' had contributed the largest share of destitutes during the famine year of 1943. Next in order of importance were 'agriculture', 'agriculture & labour', 'craft', and then 'fishing', 'profession & service' 'trade' ; least affected were 'transport', 'non-agricultural labour' and 'husking paddy'.

(7) Compared to the total population in each occupation, groups most affected as regard destitution were 'fishing' and then 'agricultural labour', 'husking paddy' 'craft', and 'transport'.

(8) The chief cause of destitution was death of earners. Next in importance was sickness and then unemployment of earners.

APPENDIX TO CHAPTER 3

3.A-1. *Age and sex distribution of the destitutes before and after the famine.* Estimated number of destitute persons in lakhs by age and sex under each class of incidence in January 1943 and May 1944, and the new destitutes created by the famine during the above period is given in table 3.A-1. Estimates of destitutes by sex and age in January 1943 and May 1944 have been derived from the proportional distribution within the sample of the destitutes surveyed. The proportional distribution of all-Bengal population by sex and age in the Census of 1931 (figures for 1941 are not available) could also have been used. But, as sex and age distribution among the destitutes may not be the same as the over-all distribution, it was considered advisable to use the former method.

TABLE 3(A-1). ESTIMATED NUMBER (IN LAKHS) OF DESTITUTE PERSONS IN EACH CLASS OF INCIDENCE BY SEX AND AGE IN JANUARY 1943 AND MAY 1944 AND THE NEW DESTITUTES CREATED BETWEEN JANUARY 1943 AND MAY 1944.

class	estimated number (in lakhs) of destitute persons											
	0—5 years			5—15 years			15—50 years			50 years and above		
	male	female	total	male	female	total	male	female	total	male	female	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
					January 1943							
Class A	0·09	0·10	0·19	0·54	0·35	0·89	0·33	0·66	0·99	0·17	0·26	0·
Class B	0·10	0·11	0·21	0·58	0·38	0·96	0·35	0·70	1·05	0·18	0·28	0·46
Class C	0·08	0·10	0·18	0·49	0·33	0·82	0·30	0·59	0·89	0·15	0·23	0·38
rural Bengal	0·27	0·31	0·58	1·61	1·06	2·67	0·98	1·95	2·93	0·50	0·77	1·27
					May 1944							
Class A	0·27	0·23	0·50	0·85	0·76	1·61	0·67	1·29	1·96	0·19	0·23	0·42
Class B	0·14	0·17	0·31	0·62	0·60	1·22	0·37	0·81	1·18	0·14	0·25	0·39
Class C	0·19	0·11	0·30	0·72	0·32	1·04	0·39	0·82	1·21	0·26	0·37	0·63
rural Bengal	0·60	0·50	1·10	2·19	1·68	3·87	1·43	2·92	4·35	0·59	0·85	1·4
					January 1943—May 1944							
Class A	0·18	0·13	0·31	0·31	0·41	0·72	0·34	0·63	0·97	0·02	–0·03	–0·01
Class B	0·04	0·06	0·10	0·04	0·22	0·26	0·02	0·11	0·13	–0·04	–0·03	–0·07
Class C	0·11	0·01	0·12	0·23	0·01	0·22	0·09	0·23	0·32	0·11	0·14	0·2
rural Bengal	0·33	0·19	0·52	0·58	0·62	1·20	0·45	0·97	1·42	0·09	0·08	0·17

3.A-2. *Post-famine destitutes.* The above discussion gives an analysis of the effect of war and famine conditions on rates of destitution in rural Bengal. From the point of view of rehabilitation it is useful to get a consolidated picture of the distribution of post-famine destitutes. This has been given in table 3.A-2 which shows the number of destitutes in May 1944 coming from respective occupational group at the beginning of 1943 by age, sex, and cause of destitution. Figures for infants belonging to age-group 0—5 years), children age-group 5—15 years and old persons (age-group 50 years and above have been shown for both sexes taken together as rehabilitation problems would not be very different for males and females. Figures for adults and adolescents (belonging to age-group 15—50 years), however, have been shown separately for males and females.

TABLE 3(A-2). ESTIMATED NUMBER (IN LAKHS) OF DESTITUTES IN MAY 1944 IN RURAL BENGAL BY AGE, SEX, CONDITION OF EARNERS AND OCCUPATIONAL GROUPS AS IN JANUARY 1943

occupational groups as in January 1943	earners or parents dead or emigrated					earners or parents invalid					earners or parents unemployed					total
	infants (0—5)	children (5—15)	adult (15—50)		old (above 50)	infants (0—5)	children (5—15)	adult (15—50)		old (above 50)	infants (0—5)	children (5—15)	adult (15-50)		old (above 50)	
			male	female				male	female				male	female		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
agriculture	0·07	0·20	0·01	0·14	0·02	0·01	0·08	0·06	0·04	0·03	0·02	0·08	0·07	0·04	0·03	0·90
agriculture & labour	0·07	0·07	0·01	0·05	0·03	0·01	0·02	0·02	0·01	0·03	0·05	0·06	0·02	0·03	0·01	0·49
agricultural labour	0·11	0·53	0·10	0·38	0·10	0·07	0·14	0·12	0·14	0·10	0·03	0·03	0·05	0·06	0·03	1·99
non-cultivating owner	0.01	0·07	0·01	0·04	0·01	—	—	—	—	0·01	0·02	0·04	0·04	0·02	—	0·27
all agriculture	0·26	0·87	0·13	0·61	0·16	0·09	0·24	0·20	0·19	0·17	0·12	0·21	0·18	0·15	0·07	3·65
fishing	—	0·06	0·01	0·04	0·01	0·01	0·02	0·01	0·01	—	0·06	0·07	0·07	0·07	0·01	0·45
craft	0·02	0·17	0·05	0·12	0·03	—	0·05	0·03	0·01	0·01	0·01	0·01	0·02	0·02	0·01	0·56
husking paddy	0·01	0·01	—	0·02	—	—	—	—	0·01	—	—	—	—	0·01	—	0·06
transport	—	0·01	—	—	—	0·01	0·01	0·01	0·01	—	—	0·01	0·01	0·01	—	0·08
trade	—	0·10	0·01	0·07	—	0·02	0·06	0·03	0·03	0·02	—	0·02	0·01	0·02	0·01	0·40
profession & service	0·02	0·10	0·04	0·05	0·02	0·01	0·04	0·03	0·04	0·01	0·01	0·03	0·01	0·02	0·01	0·44
non-agricultural labour	—	0·04	0·01	0·01	—	—	—	0·01	—	—	0·01	0·01	0·01	0·01	—	0·11
other productive occupations	—	—	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·03	0·01	0·03	—	—	—	0·14
living on charity	0·14	0·83	0·14	0·74	0·32	0·13	0·48	0·20	0·40	0·41	0·15	0·38	0·19	0·23	0·13	4·87
total	0·45	2·19	0·40	1·67	0·55	0·28	0·91	0·53	0·71	0·65	0·37	0·77	0·50	0·54	0·24	10·76

N.B. Sampling fluctuations are necessarily large for such detailed tabulations, and figures given must be used as merely indicative of general tendencies. Estimates by subdivisions were also prepared but are not printed here; sampling errors of these estimates would be naturally still larger. These estimates may, however, prove useful for administrative purposes in connection with rehabilitation work, and persons in charge of such work may obtain this information from us if they so desire.

CHAPTER 4. ECONOMIC DETERIORATION

4.1. *The effect of famine* was not only destitution on an unprecedented scale but the lowering of the economic level of a large number of people who before the famine had some assets by selling which they succeeded in escaping total destitution but whose earning power was appreciably decreased. Many families in the group 'agriculture' (owning land of their own) were forced to sell their paddy land in part or in full and were thus reduced to the class of 'agriculture & labour' or 'agricultural labour'. Similar transitions took place from other occupations of higher average earnings to those of lower average earnings. Changes of family occupation during the year 1943 would, therefore, give a good idea of the extent of lowering of economic level of rural Bengal as a consequence of the famine.

4.2. *Change of family occupations during the famine.* Table 4.1 gives the number of and proportion of families in the sample suffering change of family occupation during the period January 1943 to April 1944.

TABLE 4.1. NUMBER OF FAMILIES IN THE SAMPLE SURVEYED BY OCCUPATIONS AS IN JANUARY 1943 AND IN MAY 1944, AND NUMBER AND PROPORTION OF CHANGE OF OCCUPATION DURING THE PERIOD.

occupational groups	number in January 1943	change between January 1943 & May 1944			number in May 1944	percentage change of occupation between January 1943 & May 1944		
		increase	decrease	balance		increase	decrease	balance
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
agriculture	4875	152	412	−260	4615	3·12	8·45	− 5·33
agriculture & labour	2511	162	330	−168	2343	6·45	13·14	− 6·69
agricultural labour	2534	396	299	+ 97	2631	15·63	11·80	+ 3·83
non-cultivating owners	860	121	64	+ 57	917	11·96	6·33	+ 5·63
fishing	220	8	47	− 39	181	3·64	21·36	−17·72
craft	800	73	101	− 28	772	9·13	12·63	− 3·50
husking paddy	256	91	24	+ 67	323	35·55	9·38	+26·17
transport	116	212	25	+187	303	182·76	21·55	+161·21
trade	1087	67	330	−263	824	6·16	30·36	−24·20
profession and service	1064	96	90	+ 6	1070	9·02	8·46	+ 0·56
non-agricultural labour	134	32	23	+ 9	143	23·88	17·16	+ 6·72
other productive occupations	287	71	38	+ 33	360	21·71	11·62	+10·09
living on charity	424	364	62	+302	726	85·85	14·62	+71·23
total	15208	1845	1845	—	15208	12·13	12·13	—

Cols. (5) & (9) of the table show that all the occupational groups had suffered change during the famine. Cols. (3), (4) and (7), (8) indicate that in all occupational groups there was both increase and decrease ; in some of the groups there was an over all or net increase and in others an over-all decrease. It would be noticed that both 'agriculture' and 'agriculture & labour'—the two highest agricultural groups—had lost from 5 to 7 per cent, while 'agricultural labour' and 'non-cultivating owner' had gained appreciably. Proportionately both 'fishing,'(—17.7%) and trade (—24.2%) had lost most heavily. 'Husking paddy' (+26.2%) and 'transport' (+161.2%) had gained very appreciably as also, of course, the group 'unproductive' with an over-all increase of 71.2%.

4.3. *Measure of economic deterioration.* Shift of families from one occupational group to another does not, however, by itself mean a raising or lowering of economic level. Everything depends on from which category to which the change is made. Thus a family which finds itself in the group 'agriculture & labour' would have a lower economic status if it came from 'agriculture' and higher economic level if it had come from 'agricultural labour'. The extent and nature of the change can be appreciated only by studying the shift of families relative to their original family occupation. Tables 4.2, 4.3 and 4.4 give in greater detail the shift of families between occupational groups.

4.4. Table 4.2 shows the actual number of families surveyed together with the family occupation as in both January 1943 and May 1944. For example, 4875 families were in the group 'agriculture' in January 1943. Out of these, 4463 had retained the same occupation in May 1944 ; but 93 families had changed over to the group 'agriculture & labour', 116 to 'agricultural labour' ; and so on. Table 4.3 gives the same information in the form of weighted percentage (with total number of families surveyed=100). Finally, table 4.4 shows the estimated number (in lakhs) of families in the different categories.

4.5. *Economic deterioration in rural Bengal during the famine.* From table 4.4 it is possible to compile figures in such a way as to show whether there was any improvement or deterioration in economic status due to the change in family occupation. Table 4.5 gives the estimated total number (in lakhs) of families in each occupational group which had retained or changed their occupation between January 1943 and April 1944 classified under the three categories ; (a) improvement, (b) deterioration, and (c) ambiguous. Certain definite rules naturally had to be adopted to prepare this table. For example, a family shifting from the group of 'agriculture & labour' to 'agriculture' was considered as having improved its economic status and was thus placed in class (a) : improvement. A family originally in 'agriculture & labour' going over to 'profession and service' was classed under (b) : deterioration, as the family had sunk to a lower level in which the earner probably had to accept a job as Union Board *chowkidar* or some such thing. In certain cases it was not clear whether the change was an 'improvement' or a 'deterioration', such as, a shift from 'craft' to 'agriculture' ; these changes were classed under (c) : ambiguous. The method adopted is shown in detail in the appendix.

TABLE 4.2. INTER-OCCUPATIONAL SHIFT AMONG FAMILIES SURVEYED BETWEEN JANUARY 1943 AND MAY 1944

occupational groups in January 1943	number of families in different occupations in May 1944													
	agri-culture	agri-culture & labour	agri-cultural labour	non-cultiva-ting owners	fishing	craft	husking paddy	trans-port	trade	profe-ssion & service	non-agri-cultural labour	other produc-tive occupa-tions	living on charity	total
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
agriculture	4463	93	116	63	3	17	8	7	17	11	1	14	62	4875
agriculture & labour	65	2181	170	12	—	6	14	2	7	8	6	5	35	2511
agricultural labour	23	38	2235	1	2	12	38	4	8	30	5	21	117	2534
non-cultivating owners	5	1	7	796	—	9	7	1	4	10	2	4	14	860
fishing	1	—	5	1	173	1	2	1	10	—	2	3	21	220
craft	14	4	23	9	1	699	4	2	2	7	4	1	30	800
husking paddy	—	1	2	—	1	2	232	1	2	—	1	2	12	256
transport	1	2	9	..	..	4	1	91	..	1	..	..	7	116
trade	20	7	30	12	1	6	4	19	757	22	5	9	24	1087
profession & service	15	4	10	13	—	6	5	3	10	974	1	1	22	1064
non-agricultural labour	5	3	7	..	..	..	1	..	..	..	111	2	5	134
other productive occupations	2	6	5	2	—	2	—	—	1	5	—	289	15	327
living on charity	1	3	12	8	..	8	7	1	6	2	5	9	362	424
total	4615	2343	2631	917	181	772	323	303	824	1070	143	360	726	15208

TABLE 4.3. WEIGHTED PERCENTAGE OF INTER-OCCUPATIONAL SHIFT AMONG FAMILIES SURVEYED BETWEEN JANUARY 1943 AND MAY 1944 IN RURAL BENGAL

occupational groups as in January 1943	weighted percentage of families under respective occupations in May 1944													
	agri-cuture	agricul-ture & labour	agri-cultural labour	non-cultiva-ting owner	fishing	craft	husking paddy	transport	trade	profe-ssion & service	non-agri-cultural labour	other pro-ductive occupa-tions	living on charity	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
agriculture	29·71	0·60	0·75	0·53	0·02	0·13	0·05	0·05	0·12	0·07	0·01	0·07	0·37	32·48
agriculture & labour	0·37	14·65	1·10	0·08	—	0·04	0·10	0·01	0·04	0·04	0·03	0·02	0·19	16·67
agricultural labour	0·14	0·27	15·00	0·01	0·01	0·08	0·23	0·03	0·05	0·18	0·03	0·15	0·68	16·86
non-cultivating owner	0·03	0·01	0·05	5·71	..	0·05	0·04	0·01	0·04	0·05	0·01	0·02	0·09	6·11
fishing	0·01	—	0·03	0·01	1·08	0·01	0·01	0·01	0·06	—	0·01	0·01	0·12	1·36
craft	0·10	0·03	0·13	0·05	—	4·37	0·03	0·01	0·01	0·04	0·02	0·01	0·17	4·97
husking paddy	..	0·01	0·01	..	..	0·02	0·87	0·70	0·01	..	0·01	0·01	0·08	1·72
transport	—	0·01	0·05	—	—	0·02	0·01	0·53	—	0·01	—	—	0·03	0·66
trade	0·11	0·04	0·18	0·09	..	0·04	0·03	1·10	4.80	0·13	0·03	0·06	0·14	6·75
profession & service	0·08	0·02	0·04	0·08	—	0·03	0·03	0·01	0·06	6·11	0·01	0 01	0·13	6·61
non-agricultural labour	0·02	0·03	0·05	..	..	..	0·01	..	..	..	0·82	0·02	0·02	0·98
other productive occupations	0·01	0·03	0·03	0·01	—	0·01	—	—	—	0·03	—	1·96	0·09	2·17
living on charity	..	0·02	0·06	0·04	..	0·05	0·04	..	0·03	0·01	0·03	0·07	2·31	2·66
total	30·58	15·72	17·48	6·61	1·11	4·85	1·45	2·46	5·22	6·67	1·01	2·41	4·43	100·00

Weights used are the estimated total number of families in January 1943 in the classes A, B and C.

TABLE 4.4. ESTIMATED NUMBER (IN LAKHS) OF FAMILIES IN DIFFERENT OCCUPATIOAL GROUPS AS IN JANUARY 1943 AND AS IN MAY 1944 IN RURAL BENGAL.

occupational groups as in January 1943	estimated number (in lakhs) of families in occupational groups in May 1944													
	agri-culture	agri-culture & labour	agri-cultural labour	non-cultiva-ting owner	fishing	craft	husking paddy	trans-port	trade	prote-ssion & service	non-agri-culture labour	other produc-tive occupa-tions	living on charity	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
agriculure	30·4	0·6	0·8	0·5	—	0·1	0·1	0·1	0·1	0·1	—	0·1	0·4	33·3
agriculture & labour	0·4	15·0	1·1	0·1	..	0·1	0·1	..	..	..	0·1	..	0·2	17·1
agricultural labour	0·2	0·3	15·4	—	—	0·1	0·2	—	0·1	0·2	—	0·2	0·6	17·3
non-cultivating owner	..	..	0·1	5·8	..	0·1	..	..	..	0·1	..	..	0·1	6·2
fishing	—	—	—	—	1·1	—	—	—	0·1	—	—	—	0·1	1·3
craft	0·1	..	0·1	0·1	..	4·5	0·1	..	..	..	..	..	0·2	5·1
husking paddy	—	—	—	—	—	—	0·9	0·7	—	—	—	—	0·1	1·7
transport	..	..	0·1	..	..	..	..	0·6	..	..	..	..	..	0·7
trade	0·1	0·1	0·2	0·1	—	—	—	1·1	4·9	0·1	—	0·1	0·2	6·9
profession & service	0·1	..	..	0·1	..	..	..	..	0·1	6·3	..	..	0·2	6·8
non-agricultural labour	—	—	0·1	—	—	—	—	—	—	—	0·9	—	—	1·0
other productive occupations	..	0·1	..	..	..	..	..	..	..	..	..	2·0	0·1	2·2
iving on charity	—	—	0·1	—	—	0·1	0·1	—	—	—	—	0·1	2·4	2·8
total	31·3	16·1	18·0	6·7	1·1	5·00	1·5	2·5	5·3	6·8	1·0	2·5	4·6	102·4

TABLE 4.5. ESTIMATED TOTAL NUMBER (IN LAKHS) OF FAMILIES IN RURAL BENGAL BY FAMILY OCCUPATIONS AS IN JANUARY 1943, AND THE CHANGE BETWEEN JANUARY 1943 AND MAY 1944 CLASSIFIED UNDER 'IMPROVEMENT', 'DETERIORATION', OR 'AMBIGUOUS'.

	estimated number (in lakhs) of families				percentage of families suffering change between January 1943 & May 1944		
occupational groups	January 1943	change between January 1943 & May 1944					
		improvement	deterioration	ambiguous	improvement	deterioration	ambiguous
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
agriculture	33 ·3	—	2 ·5	—	—	7 ·51	—
agriculture & labour	17 ·1	0 ·4	1 ·5	—	2 ·34	8 ·77	—
agricultural labour	17 ·3	0 ·7	0 ·6	—	4 ·05	3 ·47	—
non-cultivating owner	6 ·2	—	0 ·2	0 ·1	—	3 ·23	1 ·61
fishing	1 ·3	—	—	0 ·1	—	—	7 ·69
craft	5 ·1	—	0 ·3	0 ·1	—	5 ·88	1 ·96
husking paddy	1 ·7	0 ·7	—	—	4 ·12	—	—
transport	·7	—	0 ·1	—	—	14 ·29	—
trade	6 ·9	—	1 ·6	0 ·2	—	23 ·19	2 ·90
profession & service	6 ·8	0 ·1	0 ·1	0 ·1	1 ·47	1 ·47	1 ·47
non-agricultural labour	1 ·0	—	0 ·1	—	—	10 ·00	—
other productive occupations	2 ·2	0 ·1	—	—	4 ·55	—	—
living on charity	2 ·8	0 ·4	—	—	14 ·29	—	—
total	102 ·4	2 ·4	7 ·0	0 ·6	2 ·34	6 ·84	0 ·59

Cols. (4) and (7) of table 4.5 show that about seven lakhs of families (6.8% of the total number in rural Bengal) had suffered a deterioration of economic status during the famine; about 2.4 lakhs (2.3% of the total) had prospered (cols. 3 and 6); and for 0.6 lakhs (0.6% of the total) it is not clear whether they had prospered or not (cols.5 and 8). Col.(4) of the table shows that deterioration was heaviest among agricultural groups. This is just what was to be expected. Most of the rural families in Bengal are engaged in agricultural pursuits, and land alienation and loss of plough cattle brought about by the famine naturally hit them hardest.

4.6. Rural 'trade' had also suffered heavily. This again was almost inevitable. Owing partly to the bringing in of various kinds of Government control on one hand, and hoarding and black marketing on the other hand, trade in rural areas could hardly function normally. There was also an appreciable deterioration in the position of families following 'crafts' which probably had its origin in the disintegration of rural industries caused by lack of raw materials and accessories. Col.(7) of the table which gives the relative proportion of impoverished families (percentage of total in each occupational group), shows that relatively 'trade' was most seriously affected and may therefore claim priority as regards rehabilitation measures.

4.7. 'Agricultural labour', however, shows certain special features. It has been n in para 3.9 that this group had suffered destitution very heavily. In table 4.5 it is seen that the shifts into better and worse groups were nearly balanced. This was probably due to the fact that owing to increased demand and rise of daily wage rates of unskilled labour (caused by war work) able-bodied agricultural labour had increasd their earnings. On the other hand, considerable numbers were unable to secure such employment and had consequently suffered a deterioration in their economic position. The comparatively large improvement (about 14%) in the 'unproductive' group is also worth noting. This also might have been brought about by the increase in employment caused by war work.

4.8. *Summary of conclusions*.

(1) About seven lakhs of families in rural Bengal had probably suffered a lowering of economic status and decrease of earning power as a consequence of the famine. Adopting 5.4 as the average size of the family in rural Bengal (as estimated from Census 1941) the total number of persons whose economic position had seriously deteriorated (and in many cases possibly reduced to semi-destitution) was about 38 lakhs.

(2) In actual numbers, the heaviest deterioration had occurred among families which originally belonged to the two occupational groups, 'agriculture' and 'agriculture & labour', followed closely by 'trade' and then by 'craft'.

(3) Impoverishment had been proportionately greatest among families following 'trade' and then among families following agricultural occupations and 'craft'.

APPENDIX. 4.(A-1). RULES ADOPTED TO DETERMINE ECONOMIC DETERIORATION

The method adopted to interpret the nature of shift of families from one occupation to another between January 1943 and May 1944 is given below.

family occupations as in January 1943 (1)	nature of shift as regards family occupations in May 1944		
	improvement (2)	deterioration (3)	ambiguous (4)
agriculture	—	all occupations	—
agriculture & labour	agriculture	all other occupations	—
agricultural labour	agriculture, agriculture & labour, craft, trade	all other occupations	—
non-cultivating owner	agriculture	all other occupations except Col. (4)	craft, trade, agriculture & labour
fishing	agriculture, agriculture & labour, craft.	all other occupations except Col. (4)	trade
craft	—	all other occupations except Col. (4)	agriculture
husking paddy	all occupations	—	—
transport	agriculture, agriculture & labour, craft, trade	all other occupations	—
trade	—	all other occupations except Col. (4)	agriculture, agriculture & labour
profession & service	agriculture	all other occupations except Col. (4)	craft, trade
non-agricultural labour	all occupations except Cols. (3) and (4)	agricultural labour	non-cultivating owner
other productive occupations	all occupations except Col. (4)	—	agricultural labour, fishing, non-agricultural labour, non-cultivating owner
living on charity	all occupations	—	—

CHAPTER 5. LAND SALE AND MORTGAGE

LAND PROBLEM BEFORE THE FAMINE

5.1. *Classification of families by paddy land.* In an agrarian country like Bengal, land is the main asset of the people. In 1943 people affected by the famine in rural areas were therefore often obliged to sell or mortgage their land to save themselves from immediate destitution or death from starvation. The land problem had already become acute in Bengal long before the famine. To grasp the full effect of the crisis, the position immediately before the famine therefore needs consideration. For present purposes rural families were classed under four groups by the amount of paddy land in their possession : (1.1) families possessing no land, (1.2) possessing 0—2 acres, (2) possessing 2—5 acres, and (3) possessing 5 acres and above. Since paddy is the main crop[1] of the country (being both a food and a money crop) classification by amount of paddy land owned will show the main features of the land position.

5.2. Groups (1.1) and (1.2) may be pooled together to form one single group (1) to represent the poorer section of the rural population. In the opinion of many economists, 2 acres of paddy land form the subsistence requirement for a family of five (the average size of family is 5.4 according to the 1941 census, and 5.2 according to the Bengal Land Revenue Commission 1940-41). It is this group which constitutes the most numerous section of rural Bengal and which suffers most in any crisis. Group (2) constitutes the middle section of the rural population, and group (3) the upper middle and richer elements.

5.3. *Ownership of paddy land.* Subdivisions in which the lowest section (possessing no paddy land or less than 2 acres) is comparatively more numerous may be considered to be at a lower economic level (and hence more vulnerable to famine conditions) than other subdivisions in which rural families possess more paddy land on an average. Table 5.1 gives the proportionate distribution of the poorest stratum (that is, of group (1) families possessing no paddy land or less than 2 acres) in surveyed subdivisions grouped according to the Industries Department classification.

TABLE 5.1. CUMULATIVE PERCENTAGE OF SUBDIVISIONS SURVEYED BY PERCENTAGE OF FAMILIES OWNING NO PADDY LAND OR BELOW 2 ACRES IN JANUARY 1943.

percentage of families owing no paddy land or below 2 acres in January 1943	cumulative percentage of sub-divisions surveyed			
	Class A	Class B	Class C	total
(1)	(2)	(3)	(4)	(5)
above 90 %	8	9	6	7
" 80 %	54	18	24	32
" 70 %	85	27	71	63
" 60 %	100	64	88	85
" 50 %	100	91	100	98
" 40 %	100	100	100	100
no. of subdivisions surveyed	13	11	17	41

---

(1) Paddy is sown on about 88% of all cultivated land in Bengal.

In this table, col. (1) shows what percentage the group (1) families form of the total, and the other columns indicate the cumulative proportion of subdivisions in the different classes. Comparing the figures in col. (2) with corresponding figures in cols. (3) and (4) it is seen that already in January 1943, before the famine, the proportion of subdivisions having higher percentages of poorer families was heavier in Class(A), the "most severely affected" area, than in the other two. For example, in the third line we find that the proportion of subdivisions having over 70% of families belonging to group (1) was 85% in Class (A) area, and only 27% in Class (B) area, and 71% in Class (C) area.

5.4. Class (B) subdivisions (which were quite severely affected by the famine) were, however, apparently better off before the famine (as judged by the present criterion) than subdivisions in Class (C). The degree of incidence had, however, followed to some extent (but not exactly) the general economic conditions prevailing before the famine. This shows that conditions in January 1943 were already pretty bad. It can be seen from col. (5) of the present table that in 35 out of 41 or 85% of the total number of subdivisions surveyed, the proportion of group (1) families was no less than 60% ; or again, in 25 out of 41 or 63% of subdivisions the proportion of group (1) families was over 70%.

5.5. *Mean acreage of paddy land per family.* Table 5.2 gives the accumulated percentage of subdivisions having different values of mean acreage of paddy land owned per family before the famine.

TABLE 5.2. CUMULATIVE PERCENTAGE OF SUBDIVISIONS SURVEYED BY MEAN ACREAGE OF PADDY LAND OWNED PER FAMILY IN JANUARY 1943.

mean acreage of paddy land owned per family in January 1943	cumulative percentage of subdivisions surveyed			
	Class A	Class B	Class C	total
(1)	(2)	(3)	(4)	(5)
above 4·0 acres	—	9	—	2
" 3·5 "	—	9	6	5
" 3·0 "	—	27	12	12
" 2·5 "	8	27	18	17
" 2·0 "	8	27	41	27
" 1·5 "	41	82	76	66
" 1·0 "	100	100	88	95
" 0·5 "	100	100	100	100
no. of subdivisions surveyed	13	11	17	41

The table shows that a gradation of the subdivisions on the basis of mean acreage of paddy land owned per family gives roughly the same kind of picture as that supplied by Table 5.1. Thus col. (2) of the table shows that in January 1943 only in 5 out of 13, or 41% of the subdivisions in Class (A), the paddy land owned was above 1.5 acre per family, and in only 1 out of 13 or 8% of the subdivisions the mean acreage was between 2.5 and 3. Cols. (3) and (4) show that conditions in subdivisions of Class (B) and of Class (C) were somewhat better ; in 82% and 76% of the subdivisions in these two respective groups the mean acreage was above 1.5 ; and in 27% and 41.0% of subdivisions it was above 2 acres. The over-all figures in col.(5)

however indicate that in only 11 out of 41 or in about one-fourth of the subdivisions the average holding of paddy land per family was above 2 acres. This shows the acute shortage of paddy land owned by the peasantry in Bengal before the famine.

5.6. *Provincial estimate of families owning paddy land before the famine.* The precarious nature of Bengal's rural economy can be seen from Table 5.3 which gives the all-Bengal estimates of families owning different amounts of paddy land.

Table 5.3· Number, proportion and estimated total number in lakhs of families owning paddy land in January 1943 in rural Bengal.

paddy land owned in January 1943 (in acres)	total number of families surveyed	percentage of total surveyed	estimated number of families (in lakhs)
(1)	(2)	(3)	(4)
no land	6145	36·2	37·4
0—2	6169	40·5	40·1
—5	2209	15·3	16·2
above 5	1117	8·0	8·7
total	15,640	100·0	102·4

(a) Provincial estimates of the proportion of families shown in col. (3) have been derived from the estimates of each subdivision and class of incidence with the total number of families in each subdivision and class in January 1943 (obtained by extrapolation of the corresponding figures of Census 1931 and 1941) used as weights.

(b) Estimated number of families as shown in col(4) are derived from the total number of families (102·4 lakhs in January 1943) and the proportion of families in each acreage group as shown in col. (3).

It will be noticed from col. (3) of the table that in rural Bengal in January 1943 more than one-third (36.2%) had no paddy land, while about two-fifths(40.5%) owned less than 2 acres. Thus more than three-fourths (76.7%) of rural families of Bengal either had no paddy land at all or less than 2 acres ; less than one-sixth (15.3%) had paddy land between 2 to 5 acres ; and only a twelvth (8.0%) had above 5 acres. It means that, adopting 2 acres of paddy land as the critical limit, only about 23% (or less than one-fourth) of the rural families of Bengal belonged to what may be called the middle and upper level of prosperity, while about 77% (or more than three-fourths) were below the subsistence level.

5.7. *Summary of conditions before the famine.* The land problem in Bengal before the famine can be summarized in the following way :

(1) Even prior to the famine in a majority of the subdivisions (26 out of 41) more than 70% of the families owned less than 2 acres of paddy land per family (Table 5.1) ; and the mean acreage owned per family was somewhere between 1.5 and 2 acres (Table 5.2).

(2) Broadly speaking, stratification of subdivisions by amount of paddy land owned per family before the famine was roughly parallel to the degree of incidence of famine conditions as shown by the classification of the Industries Department. This is, of course, just what was to be expected. Subdivisions in which there were more families with paddy land below subsistence level were naturally more vulnerable to the famine.

(3) The over-all position in Bengal was precarious. More than three-fourths of rural families had either no paddy land or owned less than 2 acres and were thus either below the subsistence level or were just on the border line (Table 5.3).

5.8. It is against this background of conditions early in 1943 that the effect of the famine on the land problem of Bengal can be properly appreciated. The factors discussed here are sales and mortgages of paddy land during the period April 1943 to March 1944 (that is, the whole of the Bengali year 1350).

LAND SALE AND MORTGAGE DURING THE FAMINE PERIOD (APRIL 1943—APRIL 1944 ; 1350 BENGALI ERA)

5.9. *Families selling paddy land during the famine* : Table 5.4. gives the estimated number of families who were forced to sell paddy land during the famine between April 1943 and April 1944.

TABLE 5.4. NUMBER, PROPORTION AND ESTIMATED TOTAL NUMBER IN LAKHS OF FAMILIES SELLING PADDY LAND BETWEEN APRIL 1943 AND APRIL 1944 (1350 BENGAL YEAR)

		sample surveyed			
class	total families in January 1943 (in lakhs)	total families in January 1943	families selling paddy land between April 1943 & April 1944	percentage of families selling paddy land	estimated total number of families selling paddy land (in lakhs)
(1)	(2)	(3)	(4)	(5)	(6)
Class A	25·93	5114	653	12·5	3·24
Class B	33·86	5994	501	9·8	3·31
Class C	42·64	4532	292	6·3	2·69
rural Bengal	102·43	15640	1446	9·0	9·24
41 subdivisions surveyed				9·1±0·27%	

(a) Percentages of families selling paddy land, as shown in Col. (5), have been derived from the proportions of each subdivision and class of incidence with the total number of families in each subdivision and class in January 1943 used as weights.

(b) Estimated total number of families selling paddy land, as shown in Col.(6), is based on the total number of families in January 1943 of Col. (2).

(c) 0·27 is the estimated standard error of the estimated mean percentage for 41 subdivisions surveyed.

Col. (5) of the table shows that the gradation of subdivisions according to intensity of incidence is well maintained. In Class (A) no less than 12.5% of families were obliged to sell land, while corresponding figures in Class (B) and Class (C) were 9.8% and 6.3% respectively. Provincial estimates in col. (6) of the table show that out of about 102.4 lakhs of rural families, 9.2 lakhs (or 9.02% of families) were forced to sell land during the famine period. It must be remembered that only about 65 lakhs of rural families had owned any paddy land at all before the famine (as shown in col. (5) of Table 5.3). Among families owning paddy land the proportion forced to sell such land during the famine year was thus more than 14%. Col. (5) shows that for the 41 subdivisions surveyed, the percentage of families selling paddy land was 9.1% with an estimated standard error of 0.27% which is quite low showing satisfactory internal consistency of the material.

5.10. *Families selling paddy land in full.* Some differentiation has to be made between families selling their paddy land in full and families selling in part. Families selling in full (specially those who were mainly dependant on agriculture) lost their sole or chief means of livelihood while families selling in part were reduced to a lower level of living. From the point of view of rehabilitation also the former group deserves special consideration. Table 5.5 gives the provincial estimates of families selling all the paddy land they had owned before the famine.

TABLE 5.5. NUMBER, PROPORTION, AND ESTIMATED TOTAL NUMBER (IN LAKHS) OF FAMILIES SELLING PADDY LAND IN FULL BETWEEN APRIL 1943 AND APRIL 1944 (1350 BENGALI YEAR)

class	total number of families in January 1943 (in lakhs)	number of families included in survey		families selling paddy land in full between April 1943 & April 1944	
		total in January, 1943	selling paddy land in full between April 1943 & April 1944	percentage	estimated total number (in lakhs)
(1)	(2)	(3)	(4)	(5)	(6)
Class A	25·93	5114	217	3·9	1·01
Class B	33·86	5994	189	3·2	1·07
Class C	42·64	4532	69	1·2	0·53
rural Bengal	102·43	15640	475	2·9	2·61

Cols. (5) and (6) of the table have been calculated in the same way as the corrsponding cols. of Table 5.4.

It can be seen from col. (6) of the table that out of 9.2 lakhs of families selling paddy land, 2.6 lakhs or a little less than one third had to sell their land in full and thus lost their only or chief source of earning. Col. (5) of the table shows that here again the gradation among subdivisions is maintained, sales in full being proportionately higher in Classes (A) and (B) than in Class (C) ; no marked difference is, however, seen between Classes (A) and (B).

5.11. *Mortgage of paddy land during the famine.* Table 5.6 gives the estimated number of families mortgaging paddy land during the famine period of April 1943--April 1944.

TABLE 5.6. NUMBER, PROPORTION AND ESTIMATED TOTAL NUMBER (IN LAKHS) OF FAMILIES MORTGAGING PADDY LAND (WITH AND WITHOUT SALE) BETWEEN APRIL 1943 AND APRIL 1944 (1350 BENGALI YEAR).

class	total number of families in January 1943 (in lakhs)	number of families included in survey			families mortgaging paddy land between April 1943 & April 1944			
		total in January 1943	mortgaging paddy land between April 1943 & April 1944		percentage		estimated total number (in lakhs)	
			with sale	without sale	with sale	without sale	with sale	without sale
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Class A	25·93	5114	312	252	6·1	4·7	1·58	1·21
Class B	33·86	5994	492	452	8·2	6·9	2·78	2·33
Class C	42·64	4532	245	221	5·4	4·9	2·30	2·09
rural Bengal	102·43	15640	1049	925	6·5	5·5	6·66	5·63
41 subdivisions surveyed					6·3±0·24			

(a) Cols. (6) & (7) of the table are derived in the same way as col. (5) of tables 5 & 6.
(b) Cols. (8) & (9) of the table are derived in the same way as col. (6) of tables 4 & 5.

Cols. (8) and (9) of the table show that the estimated total number of families mortgaging their paddy land was 6.7 lakhs (6.5% of the total), of which 5.63 lakhs (5.5% of the total) had not sold their paddy land. Thus it is seen from col. (6) of Table 5.4 and col. (9) of table 5.6 that 14.9 lakhs of families (out of a total of 102.4 lakhs of rural families, and out of 65 lakhs of families who had originally owned paddy land) were obliged to have recourse to either sale or mortgaging of paddy land. The number of families who both sold and mortgaged was, of course, less, namely 6.7 lakhs. Col. (6) of the table also shows that the percentage of families mortgaging paddy land for the 41 subdivisions surveyed was 6.3% (together with an estimated standard error of ± 0.24% only which is again quite low).

5.12. It is seen from cols. (6)—(9) of Table 5.6, that the number of mortgages was proportionately heavier in Class (B), having been 8.2% as against 6.1% in Class (A) and 5.4% in Class (C). This probably indicates that incidence of famine having been more acute in Class (A) many families were obliged to sell their land outright instead of mortgaging them which would have fetched only about half the cash price received from sales. In Class (C) the incidence of mortgaging was lower probably because adverse conditions were less acute.

5.13. *Summary of conclusions.* The extent of land sales and mortgages during the famine period of April 1943—April 1944 can be summed up as follows :

(1) 9.2 lakhs of families of rural Bengal sold their paddy land.
(2) 2.6 lakhs of families sold their paddy land in full and thus lost their sole or chief means of livelihood.
(3) 6.7 lakhs of families of rural Bengal mortgaged their paddy land of which 1.03 lakhs had also sold a part of their holding.
(4) 14.9 lakhs of families (22.9%of the number owning paddy land) were thus obliged either to sell or mortgage their paddy land.

LAND SALES IN RELATION TO ACREAGE OWNED

5.14. Estimates of total sale of land (discussed in the previous section) supply an overall picture, but do not indicate what sections of the population had suffered most. It is known that the hardship imposed by the famine had not been uniform in different sections of of the rural population. Certain groups were reduced to destitution while others remained immune and still others had even prospered. For the planning of rehabilitation such differential incidence also have to be taken into consideration. To study this point, families were classified by the amount of paddy land owned before the famine in March 1943.

5.15. *Sales in full.* Table 5.7 gives the proportion of families selling paddy land in full in different groups of families classified by the amount of paddy land owned before the famine, viz, 0—2 acres, 2—5 acres and above 5 acres.

TABLE 5.7. NUMBER AND PROPORTION OF FAMILIES SELLING PADDY LAND IN FULL BETWEEN APRIL 1943 AND APRIL 1944 BY GROUPS OF FAMILIES OWNING DIFFERENT AMOUNTS OF PADDY LAND BEFORE THE FAMINE (IN JANUARY 1943)

class	families owning 0—2 acres, 2—5 acres and above 5 acres of paddy land								
	total number of families surveyed			families selling paddy land in full					
				total number			percentage		
	0—2	2—5	above 5	0—2	2—5	above 5	0—2	2—5	above 5
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Class A	2327	625	297	213	2	2	8·5	—	—
Class B	2096	884	443	164	20	5	8·0	1·7	1·2
Class C	1746	700	377	58	10	1	3·0	1·2	—
rural Bengal	6169	2209	1117	435	32	8	6·1	1·2	0·4

Cols. (8)-(10) of the table have been derived in the same way as col. (5) of the Tables 5.4 and 5. 5 and cols. (6) and (7) cf Table 5.6.

Col. (8) of the table shows that sales in full were most numerous 6.1% of total in the group owning 0—2 acres ; the proportion was much less in the other two groups, only 1.2% among families owning 2—5 acres and 0.4% among families owning above 5 acres as shown in cols. (9) and (10) of the table. It will be noticed from cols. (8)—(10) of the table that the Industries Department gradation was fully maintained in the 0—2 acre group. Fluctuations however occured in the other two groups of families owning from 2—5 acres or above 5 acres of paddy land, and were thus above the subsistence level. Some of these families might have sold their paddy land for reasons other than the famine. For example, land prices had increased enormously and offered obvious inducement to sales.

5.16. Table 5.8 gives the estimated total number of families in rural Bengal selling paddy land in full in different acreage groups owned before the famine.

TABLE 5.8. ESTIMATED TOTAL NUMBER (IN LAKHS) OF FAMILIES SELLING PADDY LAND IN FULL BETWEEN APRIL 1943 AND APRIL 1944 BY GROUPS OF FAMILIES OWNING DIFFERENT AMOUNTS OF PADDY LAND BEFORE THE FAMINE (IN JANUARY 1943)

paddy land owned in acres per family before famine in January 1943	estimated total number (in lakhs) of families as in col (4) of table 5.3	number of families included in survey		families selling paddy land in full	
		total	selling paddy land in full	percentage	estimated total number in lakhs
(1)	(2)	(3)	(4)	(5)	(6)
No land	37·4	6145	—	—	—
0—2	40·1	6169	435	6·1	2·43
2—5	16·2	2209	32	1·2	0·20
above 5	8·8	1117	8	0·4	0·04
total	102·4	15640	475	2·9	2·67

(a) Col (5) is derived in the same way as cols. (10)—(13) of table 5.7.
(b) Col. (6) is derived from col.(2) and col.(5) of the table.5.8.

Col. (6) of the table shows that proportionately the largest number of families (2.43 lakhs) had sold paddy land in full from among families who owned from 0—2 acres. Among families owning 2—5 acre of paddy land, the number selling in full was very small, only 20,000 ; and among families owning above 5 acres the number selling in full was negligible, only 4,000. Thus Tables 5.7 and 5.8 show that, both proportionately as well as in actual numbers, the 0—2 acre group was worst affected by the famine and must claim priority in regard to rehabilitation measures.

5.17. *Sales in part.* In contrast, part sale of paddy land was heavy among practically all acreage groups as can be seen from Tables 5.9 and 5.10 which are similar to Tables 5.7 and 5.8. Table 5.9 gives the proportion of families selling paddy land in part in different acreage groups owned before the famine.

TABLE 5.9. NUMBER AND PROPORTION OF FAMILIES SELLING PADDY LAND IN PART BETWEEN APRIL 1943 AND APRIL 1944 BY GROUPS OF FAMILIES OWNING DIFFERENT AMOUNT OF PADDY LAND BEFORE THE FAMINE (IN JANUARY 1943)

class	families owning 0—2 acres, 2—5 acres and above 5 acres of paddy land								
	total number of families surveyed			families selling paddy land in part					
				total number			percentage		
	0—2	2—5	above 5	0—2	2—5	above 5	0—2	2—5	above 5
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
class A	2327	625	297	183	178	75	8·1	27·9	24·5
class B	2096	884	443	146	120	46	8·6	14·4	12·1
class C	1746	700	377	92	91	40	5·1	13·7	10·9
rural Bengal	6169	2209	1117	421	389	161	7·4	15·3	12·5

Cols. (8)—(10) of the table have been derived in the same way as the corresponding cols. of Table 5

For the province as a whole, cols.(8)—(10) of the table show that part sales were most numerous among the middle group of families owning 2—5 acres (15.3% of the total number). Next in order was the group owning above 5 acres (12.5%) while the smallest number of part sales occured among families owning 0—2 acres (7.4%). The comparatively smaller proportion of part sales in the 0—2 acre group is due to the fact that when a poor peasant is obliged to sell his paddy land he has to do this in full, and is thus unable to retain a portion for his own use.

5.18. Coming to different classes of subdivisions, it will also be noticed from cols.(8)—(10) that the adopted gradation is maintained quite well among families owning 0—2 acres of paddy land showing that in this group of families subdivisions belonging to Classes (A) & (B) were most severely affected (8·1% and 8·6% respectively); and Class (C) was last (5.1%). Among families owning 2—5 acres also the gradation was quite clear with 27.9% of part sales in Class(A), but there was little difference between Class (B) (14.4%) and Class (C) (13.7%). Among families owning above 5 acres the Industries Department gradation was fully maintained.

5.19. Table 5.10 gives the estimated total number of families in rural Bengal selling paddy land in part in different acreage groups owned before the famine.

TABLE 5.10. ESTIMATED TOTAL NUMBER (IN LAKHS) OF FAMILIES SELLING PADDY LAND IN PART BETWEEN APRIL 1943 AND APRIL 1944 BY GROUPS OF FAMILIES OWNING DIFFERENT AMOUNTS OF PADDY LAND BEFORE THE FAMINE (IN JANUARY 1943)

paddy and owned in acres per family before famine in January 1943	estimated total number (in lakhs) of families in col (4) of table 5.3	number of families included in survey		families selling paddy land in part	
		total	selling paddy land in part	percentage	estimated total number in lakhs
(1)	(2)	(3)	(4)	(5)	(6)
No land	37·4	6145	—	—	—
0—2	40·1	6169	421	7·4	2·99
2—5	16·2	2209	389	15·3	2·49
above 5	8·8	1117	161	12·5	·10
total	102·4	15640	971	6·1	6·58

Cols (5) and (6) of the table are derived in the same way as the corresponding columns of table 5.8.

Col. (6) of the table shows that in actual numbers families owning 0—2 acres and 2—5 acres were almost equally important with 2.99 lakhs out of 40.08 lakhs in the former and 2·49 lakhs out of 16.22 lakhs in the latter group. This means that although proportionately part sales of paddy land were small in the 0—2 acre group, in actual numbers this group suffered more in this respect and hence requires rehabilitation measures as much as families in the next higher group of 2—5 acres. Among 8.7 lakhs of the upper group of families owning above 5 acres about 1.1 lakhs of families sold a portion of their paddy lands. As already pointed out, this may have been due partly to distress caused by the famine and partly to the desire to sell when prices were high.

5.20. *Summary of conclusions.* Points regarding sales of paddy land by families owning different amounts of such paddy land are summed up below:

(1) Sales in full were important only in the group owning less than 2 acres; 2.4 lakhs of families out of 40 lakhs were obliged to do this and they lost their sole or chief means

of livelihood. Looked at from a different point of view, it would be seen that out of a total of 2.6 lakhs of families selling paddy land in full no less than 2.4 lakhs belonged to the lowest economic level. Pauperisation on a large scale had evidently taken place in this stratum.

(2) Sales in part, on the other hand, had occured in large numbers in all groups. Such sales were proportionately most frequent among the middle group of families owning 2—5 acres (15.3%) and least among the lowest group owning 0—2 acres (7.4%).

(3) Evidently, in the lowest stratum, families who sold paddy land were obliged to do so in full rather than in part. In the middle group, on the other hand, many could manage to retain a portion for their own use. In actual numbers both groups, however, were heavily affected in as much as about 2.99 lakhs in the lower and 2.49 lakhs in the middle group had actually sold a part of their paddy land. There were part sales also by about 1.1 lakhs out of 8.7 lakhs (or 12.5%) of families in the upper group owning above 5 acres of paddy land. This may have been due partly to the famine and partly to the inducement of high prices of paddy land which was a characteristic feature during the famine year.

LAND SALES AND MORTGAGES IN RELATION TO OCCUPATIONAL GROUPS BEFORE THE FAMINE

5.21. It is of some importance to consider the incidence of land sales and mortgages among families in different occupational groups. Sale of paddy land may mean impoverishment for a non-agricultural family having other means of livelihood ; but, for an agricultural family in the lowest stratum, sale of paddy land may mean destitution. From the point of view of rehabilitation it is the different groups under the broad occupational category of Agriculture (as defined in para 2.20) which are of the greatest importance. It is these groups ('agriculture', 'agriculture & labour', 'agricultural labour', and 'non-cultivating owner') which have been studied in detail in the present section.

5.22. *Families owning paddy land in each occupational group.* For a proper appreciation of the distress caused in different occupational groups it is necessary to have as a background the distribution of paddy land among these groups before the famine. Table 5.11 gives the proportion of families (in the sample) owning paddy land in APRIL 1943 by family occupation before the famine (i.e. in January 1943).

TABLE 5.11. NUMBER AND PROPORTION OF FAMILIES SURVEYED OWNING PADDY LAND IN APRIL 1943 BY FAMILY OCCUPATIONS BEFORE THE FAMINE IN JANUARY 1943.

occupational groups before the famine in January 1943	number of families included in survey		percentage of families owning paddy land to total
	total	owning paddy land	
(1)	(2)	(3)	(4)
agriculture	5090	4285	84·2
agriculture & labour	2616	1918	73·3
agricultural labour	2506	395	15·8
non-cultivating owner	922	922	100·0
others	4506	1737	38·5

Col. (4) of the table shows that all the families in the group 'non-cultivating owner' had paddy land before the famine because such ownership was itself the basis of this particular classification. Excluding this group of 'non-cultivating owner', it will be seen from the same column, that the largest proportion of families owning paddy land (84%) belonged to the group 'agriculture'. This again is as it should be, since, by definition, this group consisted of families which were wholly or predominantly dependant on cultivation ; the remaining 16% (who did not own paddy land) did their cultivation on the basis of share cropping.

5.23. The proportion of families owning paddy land in the group 'agriculture & labour' (73%) was also large in conformity with the definition of this group. This section comes immediately after agriculturists proper in the economic and social scale. Not owning sufficient land nor getting enough land to cultivate on *barga*, these families are forced to work as hired labourers as well, but their chief source of livelihood is cultivation.

5.24. In the group 'agricultural labour' (16%) the proportion of families owning paddy land was naturally very small. Among 'others' (i.e. non-agricultural families) the proportion (39%) owning paddy land was also small. Thus for 'agricultural labour' and 'others' land transfer or mortgage during the famine was not an important problem as these families did not own any land or so little that it was only a minor source of income to them.

5.25. *Land transfer and mortgage in each occupational group during the famine.* Table 5.12 gives the proportion of families in each occupational group owning paddy land before the famine who were obliged to sell or mortgage their paddy land during the famine.

TABLE 5.12. NUMBER AND PROPORTION OF FAMILIES MORTGAGING OR SELLING PADDY LAND (IN FULL AND IN PART) BETWEEN APRIL 1943 AND APRIL 1944 BY FAMILY OCCUPATIONS BEFORE THE FAMINE IN JANUARY 1943.

occupational groups before famine in January 1943	number of families surveyed				percentage of families		
	total	selling paddy land		mortgaging paddy land	selling paddy land		mortgaging paddy land
		in full	in part		in full	in part	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
agriculture	5090	167	487	364	2·36	9·90	7·15
agriculture & labour	2616	141	196	198	5·96	6·69	7·57
agricultural labour	2506	49	52	60	2·08	2·04	2·39
non-cultivating owner	922	21	88	44	2·17	4·88	4·77
others	4506	85	148	131	2·51	2·55	2·91

Cols. (6)—(8) of the table have been derived in the same way as Col. (5) of Table (5,4).

Col. (6) of the table shows that sales in full occured most heavily in the 'agriculture & labour' group, namely in about 6% of all families in this group, while in the other occupational groups the proportion was of the order of only 2% . Proportion of part sales, shown in col. (7), was highest in the group 'agriculture' (9.9%) ; moderately high in 'agriculture & labour' (6.7%) and 'non-cultivating owner' (4.9%); and least in the other two groups each with about 2% only. Col. (8) of the table shows that the proportion of mortgage of paddy land was heaviest in the two groups, 'agriculture' and 'agriculture & labour', with 7.2% and 7.6% respectively ; it was appreciable in the group 'non-cultivating owner' (4.8%) ; but comparatively small in the remaining two groups 'agricultural labour' and 'others', (2.4% and 2.9% only).

5.26. *Summary of conclusions.* The incidence of land sales and mortgages during the famine by occupational groups (i.e. among the agricultural groups and 'others') is summarized below.

(1) Out of the five occupational groups under consideration, the groups 'agriculture' (85.2%), 'agriculture & labour' (74.2%) and 'non-cultivating owner' (100%) naturally

had the highest proportions of families owning paddy land before the famine. Proportion of families owning paddy land in the remaining two groups 'agricultural labour' (16.0%) and 'others' (38.9%) were naturally smaller.

(2) Proportion of families either mortgaging or selling paddy land in part was heaviest (7.2% and 9.9% respectively) in the group 'agriculture', but sales in full were comparatively low (2.4%). This indicates that due to the famine there was impoverishment among many families in the group but not pauperisation on a large-scale. Also some of the better-off families in this group had probably sold paddy land in part as prices had increased enormously.

(3) The group 'agriculture & labour' was most severely affected in all respects (6.0% of sales in full, 6.7% of sales in part and 7.6% of mortgages) and requires high priority in rehabilitation measures. Like 'agriculture' this group depends mainly on cultivation but owns less land than the former and is thus situated lower in the socio-economic scale. The result has been that the famine not only caused more widespread and greater distress among families in this group but also more seriously affected their future. In the 'agriculture' group there was impoverishment but in the 'agriculture & labour' group there was pauperisation on a large scale.

(4) Some of the families in the 'non-cultivating owner' group were also obliged to sell or mortgage their paddy land. Probably these were families of widows or invalids who did not own much land but who belonged to this group because they had to get their land cultivated by others. The number of families in this group is small (only 5.9% of the total in the sample); and the majority is well off, so that, proportions of sales in full as well as sales in part or mortgages were low. They do not therefore present a serious problem for rehabilitation.

(5) So far as land transfer and mortgage are concerned, the groups 'agricultural labour' and 'others' also do not present any serious problem simply because proportions of families owning paddy land were small even before the famine in these two groups. But here also an appreciable numbers of sales in full or in part or mortgages indicate widespread distress and disintegration of rural economy.

LAND PROBLEM AFTER THE FAMINE

5.27. The land problem that faces the country at present, after the famine, may now be formulated from the figures discussed in previous sections. Even before the famine, shortage of paddy land was acute in Bengal. In the province as a whole about a third of all rural families did not own any paddy land while two-fifths had less than 2 acres, so that about three-fourths of all families had no paddy land or less than 2 acres. Subdivisions in which families owned, on an average, less land were naturally more severely affected by the famine. Families holding less land were obliged to sell out more heavily which has further increased the number in the lower economic levels. Mortgaging was also comparatively heavy among the less favourably placed families. Much of this land is not likely to be redeemed and would gradually pass into other hands leading to a further increase in the proportion of families owning little or no paddy land. During the famine, 2.6 lakhs of families (out of 65 lakhs owning paddy land) had totally lost their holdings, and were thus reduced to the rank of landless labour. This has made the agricultural economy still more top heavy.

5.28. Even before the famine, several groups had become differentiated within the general category of agricultural occupations which have been designated in this report as 'agriculture',

'agriculture & labour', 'agricultural labour' and 'non-cultivating owner'. The famine has caused the transfer of an appreciable number of families from the group of 'agriculture' to 'agriculture & labour', and from 'agriculture & labour to 'agricultural labour'; and has thus accentuated inequalities in the distribution of paddy land in the province. This is, of course, not surprising and could have been anticipated on *a priori* grounds. The present sample survey has, however, supplied a factual basis and has shown the extent and direction of changes caused by the famine in concrete terms.

5.29. Accentuation of the maldistribution of paddy land in the province is finally confirmed by Table 5.13 which gives the total acreage bought and sold by the agricultural groups between April 1943 and April 1944.

TABLE 5.13. ESTIMATED ACREAGE OF PADDY LAND (IN LAKHS) TRANSFERRED BETWEEN APRIL 1943 AND MARCH 1944 BY OCCUPATIONAL GROUPS AS IN JANUARY 1943.

occupational groups as in January 1943	estimated acreage of paddy land transferred in lakhs of acres between April 1943 & April 1944			percentage of bought to sold
	bought	sold	difference	
(1)	(2)	(3)	(4)	(5)
agriculture	1·74	3·06	−1·32	57·5
agriculture & labour	0·17	1·45	−1·28	11·7
agricultural labour	0·05	0·51	−0·45	9·8
non-cultivating owner	0·38	0·77	−0·39	49·4
others	0·53	1·31	−0·78	40·5
total	2·87	7·11	−4·24	40·4

Col. (5) gives col. (2) as percentage of col. (3).

Col. (4) of the table shows that roughly 4.24 lakhs acres of land (subject to errors of sampling) which had been sold in the villages were not bought by families within the villages. Ownership of such lands has thus passed to outsiders possibly into the hands of non-cultivating owners residing in urban areas, and has added to the inequality of distribution. Furthermore, col. (5) (which shows land purchased as percentags of land sold) indicates that about half of the land sold by 'agriculture', 'non-cultivating owner' and 'other' families had passed back into the same occupational groups through purchase. But in the case of 'agriculture & labour' and 'agricultural labour' families only about 10% of the land sold had been replaced by purchase showing that these two classes had suffered serious net loss of paddy land.

CHAPTER 6. LOSS OF PLOUGH CATTLE.

6.1. *Plough cattle in Bengal agriculture.* Next to land, plough cattle is the main asset of rural families in Bengal. The extent of loss of plough cattle would thus not only indicate the severity of famine conditions but would also show an important aspect of the disintegration of rural economy. As bullocks supply the chief animal labour in agriculture in Bengal, plough cattle would usually mean bullocks.

6.2. *Cattle position before the famine.* According to the official cattle census the total number of plough cattle in rural Bengal was 86.99 lakhs in 1940 and 99.89 lakhs in 1930. The trend between 1930 and 1940 being one of decrease, the total number for January 1943 has been extrapolated as 83.12 lakhs. Table 6.1 gives the estimated total number of plough cattle in rural Bengal obtained from the sample survey.

TABLE 6.1. ESTIMATED TOTAL NUMBER OF PLOUGH CATTLE IN RURAL BENGAL IN APRIL 1943.

class (1)	total number (in lakhs) of families in January 1943 (2)	mean number of plough cattle per family in April 1943 (3)	estimated total number (in lakhs) of plough cattle in April 1943 (4)
Class A	25 ·9	0 ·58	15 ·02
Class B	33 ·9	0 ·78	26 ·44
Class C	42 ·6	0 ·88	37 ·49
rural Bengal	102 ·4	0 ·77	78 ·95

Col. (4) of the table shows that the estimated total number is 79 lakhs. The agreement between the sample estimate and the value (83 lakhs) extrapolated from census figures is quite satisfactory. As the extrapolated figure is subject to unknown errors of estimation it has been considered advisable to adopt the sample figure as the standard, but in certain cases both census and sample estimates have been used in calculations.

6.3. *Cattle shortage before the famine.* It is difficult, for reasons explained in paragraph 2.22 to quote a definite figure for the area under paddy. The present enquiry supplied a sample estimate of 186 lakhs of acre ; the official figure for area under *aman* paddy in 1943-44 was 182 lakhs of acre ; the crop survey estimate was appreciably in excess of 200 lakhs of acre. The acre per pair of bullocks thus comes out as something of the order of 5 acres on an average. The area that can be tilled varies naturally from district to district depending on the nature of the soil, and also from family to family, depending on the kind of cattle used and the nature and amount of food given to the animals. The usually accepted estimates of the area which can be tilled by a pair of bullocks varies from 4 to 5 acres according to many economists. It may be said, therefore, that the total number of plough cattle in Bengal before the famine was just sufficient or fell slightly short of requirements.

6.4. *Cattle position after the famine.* Table 6.2 incidates the position of plough cattle, after the famine, based on estimates of the net loss sustained between April 1943 and April 1944 (that is, during the Bengali year 1350).

TABLE 6.2. NUMBER, PROPORTION AND THE ESTIMATED NUMBER (IN LAKHS) OF PLOUGH CATTLE IN APRIL 1944 AND THE NET LOSS BETWEEN APRIL 1943 AND APRIL 1944.

class	number of plough cattle (sample surveyed)			percentage loss (April 1943 =100)	estimated number in lakhs of plough cattle					
					April 1943		April 1944		net loss	
	April 1943	April 1944	net loss		census 1940	sample survey	census 1940	sample survey	census 1940	sample survey
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
class A	2764	2318	446	16·14	16·32	15·13	13·69	12·69	2·63	2·44
class B	4561	3778	783	17·17	30·48	26·46	25·25	21·92	5·23	4·54
class C	4085	3738	347	8·50	36·32	37·29	33·23	34·12	3·09	3·17
rural Bengal	11410	9834	1576	13·18	83·12	78·88	72·17	68·73	10·95	10·15
41 subdivisions surveyed				13·81						

Weights used are proportional to the number of plough cattle (according to the Census of 1940) in each class of the Industries' Department classification.

The loss of cattle has been calculated on the basis of the cattle census [cols.(6), (8), and (10)] and also of the sample survey [cols. (7), (9) and (11)]. It is seen that the sample basis gives a total loss of about 10 lakhs (col.11) against about 11 lakhs on the census basis (col.10). The number available after the famine is thus 69 lakhs on the sample and 72 lakhs on the census basis. In round figures there were about 70 lakhs or a little more of plough cattle in April 1944. In any case a loss of 13% of plough cattle must have seriously aggravated the shortage.

6.5. *Cattle loss during the famine.* It will be seen from col.(5) of Table 6.2 that the proportion of net loss of plough cattle does not tally with the classification, the proportion being heavier, 17.2% in Class (B) as compared to 16.1% in Class (A). Loss of cattle may, however, occur in two different ways: by death or by sale (due to distress). A more detailed analysis is therefore necessary. Table 6.3 gives the proportion and estimated number of cattle lost during the famine through 'death' or 'sale', and also the proportion and number purchased.

TABLE 6.3. PROPORTION AND ESTIMATED NUMBER OF PLOUGH CATTLE LOST THROUGH DEATH OR SALE AND GAIN BY PURCHASE BETWEEN APRIL 1943 AND APRIL 1944.

class	number of plough cattle (sample surveyed)				percentage change (sample surveyed)			estimated number in lakhs of plough cattle			
	owned in April 1943	change between April 1943 & April 1944						owned in April 1943 (census figure)	change between April 1943 & April 1944		
		dead	sold	pur-chasd	dead	sold	pur-chased		dead	sold	pur-chased
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Class A	2764	152	400	106	5·5	14·5	3·8	16·32	0·90	2·37	0·62
Class B	4561	457	604	278	10·0	13·2	6·1	30·48	3·05	4·02	1·86
Class C	4085	124	335	112	3·0	8·2	2·7	36·32	1·09	2·98	0·98
rural Bengal	11410	733	1339	496	6·1	11·3	4·2	83·12	5·04	9·37	3·46

Weights used are proportional to the number of plough cattle (according to the Census of 1941) in each class of the Industries' Department classification.

Col.(7) of the table shows that as regards sales of plough cattle the gradation by class is fully maintained. The immediate effect of distress would be to stimulate sales of cattle ; it would be seen, in fact, that the effect of the famine was more severe in this regard in Class (A). Cols.(6) and (8), however, indicate that both the rate of death and the rate of purchase were heavier in Class (B) than in Class (A). The heavier death rate is somewhat anomalous, but may have been partly due to the outbreak of epidemics known to have occurred in certain (B) areas, such as, Bhola in Bakherganj district and Nilphamari in Rangpur district.

6.6. Cols.(6)—(8) and (10)—(12) show that in all classes 'sales' were heavier than 'deaths', and were nearly twice as frequent as deaths for the province as a whole. Replacement was comparatively small being only a third or a fourth of the loss by sales. It should be noted that the difference between sales and purchases was much greater than any thing that may be considered to have arisen from *sampling fluctuations*. It is clear, therefore, that the transfer of cattle did not take place merely from one rural family to another, but that there were large purchases by outsiders (possibly military contractors for the supply of meat for consumption by the army) as movements of cattle from Bengal to other provinces were negligibly small.

6.7. Provincial estimates, as shown in cols. (10)—(12) of Table 6.3, indicate that of the total gross number lost (about 14.4 lakhs), sales accounted for 65.0% (9.4 lakhs) and deaths for 35.0% (5.0 lakhs). Purchases made up for the loss to the extent of only 24.0% (3.5 lakhs).

6.8. *Family distribution of loss of cattle before and after the famine.* Figures for the over-all loss of cattle does not, however, show how the families which owned plough cattle were affected. It is clear that if an agricultural family loses all its plough cattle then it would become very seriously handicapped for agriculture. On the other hand, a family losing a portion of its cattle might be still able to do something. The position would be, therefore, far more serious if a considerable proportion of families had lost all their cattle than if the total loss had been shared partially by most of the families. Table 6.4 gives the proportion and estimated total number of families which owned plough cattle before the famine in April 1943 and after the famine in April 1944.

Table 6.4. Number, proportion and estimated number (in lakhs) of families owning plough cattle in April 1943 and April 1944.

	number surveyed			percentage of total families owning plough cattle		estimated total no. in lakhs of families			
class	total	owning plough cattle				January 1943	owning plough cattle		losing all plough cattle between April 1943 & April 1944
		April 1943	April 1944	April 1943	April 1944		April 1943	April 1944	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Class A	4731	1267	1114	26·78	23·54	25·9	6·94	6·10	0·84
Class B	5844	2001	1767	34·24	30·24	33·9	11·61	10·25	1·36
Class C	4667	1924	1830	41·22	39·21	42·6	17·56	16·70	0·86
rural Bengal	15242	5192	4711	35·26	32·28	102·4	36·11	33·05	3·06

Weights used for preparing the provincial estimates are the number of families (according to the Census of 1941) in each class of the Industries Department classification.

Col.(10) of the table shows that in rural Bengal as a whole, 3.06 lakhs of families (or 8.5% out of 36.11 lakhs of families which had owned plough cattle in April 1943) had lost all the cattle they had. These families were, therefore, practically incapacitated for agricultural operations.

6.9. *Summary of conclusions.*

(1) The total number of plough cattle in Bengal before the famine was just sufficient or slightly less than adequate for the cultivation of *aman* paddy, the main crop of the province.

(2) The estimated net loss of plough cattle in Bengal was from ten to eleven lakhs (about 13%) during the famine period between April 1943 and April 1944.

(3) This heavy loss of plough cattle during the famine is likely to affect seriously agricultural operations in Bengal in future.

(4) Sales (9.4 lakhs) accounted for about 65% or the greater part of the loss of plough cattle; and cattle deaths (5.0 lakhs) for about 35% of the loss. Replacement by purchases (3.5 lakhs) was to the extent of only about 24% of the total loss.

(5) Sales of cattle largely exceeded purchases showing that transfers were made not merely from one rural family to another, and that large purchases were made by outsiders (possibly by contractors for the supply of meat for army consumption).

(6) About 3.06 lakhs or 8.5% of all families of rural Bengal had probably lost all the cattle they had before the famine making it difficult if not practically impossible for them to carry on normal agricultural operations. This must have a great deal of adverse effect on the rural economy of the province.

PLOUGH CATTLE AMONG AGRICULTURAL GROUPS

6.10. *Position before the famine.* Since plough cattle is employed mainly for the cultivation of paddy, the distribution of plough cattle in relation to paddy land owned per family (in the sample) would give a proper idea of the real position. Such distribution among different agricultural groups as in April 1943 is shown in Table 6.5.

TABLE 6.5. NUMBER AND PROPORTION OF PLOUGH CATTLE AND AREA OF PADDY LAND OWNED PER FAMILY IN APRIL 1943 IN DIFFERENT OCCUPATIONAL GROUPS IN SAMPLE SURVEYED

occupational groups in 1943	plough cattle		paddy land		average per family		share of paddy land owned per pair of cattle
	total families	number owned	total families	acreage owned	plough cattle	paddy land	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
agriculture	4852	7614	5090	14710	1·57	2·89	3·7
agriculture & labour	2425	1875	2616	2589	0·77	0·99	2·6
agricultural labour	2598	332	2506	400	0·13	0·16	2·5
non-cultivating owner	886	627	922	3309	0·71	3·59	10·1
others	4481	962	4506	2613	0·21	0·58	5·5

Col.(8) of the table shows that in the group 'agriculture' the share of paddy land owned per pair of cattle was 3.7 acres before the famine. This means that in most cases these families failed to utilize their plough cattle to the fullest extent unless they secured enough additional land on share cropping basis. It will be seen from the same column that the mean acreage per pair of cattle in the groups 'agriculture & labour' and 'agricultural labour' also were small namely 2.6 and 2.5 acres respectively. Because of the small amount of paddy land owned per family in these two groups, only a very small number could properly work their cattle on their own land. It will also be seen that among 'non-cultivating owner' the share per pair of cattle was 10 acres of paddy land which is quite natural as, by definition, these families had their land cultivated principally on a *barga* (share) basis where the *bargadar* (share-cropper) supplied the cattle. Thus, there was possibly an over-all shortage of cattle and yet cultivating groups had more cattle than they could fully employ on their own lands. The problem of plough cattle, like that of land, was thus complicated by anomalies of distribution even before the famine.

6.11. *Loss of cattle during the famine.* Table 6.6 which gives the loss of plough cattle by occupational groups during the famine shows that such anomalies were further aggravated by the famine.

TABLE 6.6. NUMBER AND PROPORTION OF PLOUGH CATTLE LOST BETWEEN APRIL 1943 AND APRIL 1944 BY OCCUPATIONAL GROUPS

occupational groups in 1943	owned in April 1943	percentage change between April 1943 and April 1944			
		died	sold	purchased	net loss
(1)	(2)	(3)	(4)	(5)	(6)
agriculture	7614	4·5	9·6	3·1	11·0
agriculture & labour	1875	8·0	15·1	5·5	17·6
agricultural labour	332	9·0	28·8	8·5	29·3
non-cultivating owner	627	13·6	8·9	7·7	14·8
others	962	8·5	14·3	5·9	16·9

Col.(6) of the table shows that all occupational groups had lost cattle, but the net loss was heaviest in the group 'agricultural labour', next in 'agriculture & labour' and then in 'others.' From col.(4) it is seen that loss due to sales (which supplies a good indication of the severity of distress) was also highest in the group 'agricultural labour', next in 'agriculture & labour' and then in 'others.' It is interesting to note, however, that loss due to deaths was highest in the group of 'non-cultivating owner' and lowest in 'agriculture'.

6.12. *Cattle position after the famine.* Table 6.7 which gives the mean number of plough cattle owned per family in April 1943 and in April 1944 shows how the famine affected the cattle position in each occupational group.

TABLE 6.7. NUMBER OF PLOUGH CATTLE OWNED PER FAMILY IN APRIL 1943 AND IN APRIL 1944 BY OCCUPATIONAL GROUPS.

occupational groups in 1943	plough cattle in April 1943		plough cattle in April 1944		mean number of plough cattle owned		change in mean number between April 1943 and April 1944
	owning families	number owned	owning families	number owned	April, 1943	April 1944	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
agriculture	3288	7614	3077	6745	2·32	2·19	−0·13
agriculture & labour	988	1875	863	1537	1·90	1·78	−0·12
agricultural labour	179	332	128	226	1·85	1·76	−0·09
non-cultivating owner	252	627	228	543	2·49	2·38	−0·11
others	485	962	415	783	1·98	1·88	−0·10

Col.(8) of the table shows that in each occupational group there was some decrease in the average number of plough cattle owned per family after the famine, but cols.(6) and (7) indicate that families, on the whole, still tended to own at least a pair, for the obvious reason that a pair is indispensable for cultivation purposes.

6.13. *Families owning plough cattle before and after the famine.* Consequences of loss of cattle can be fully appreciated only when comparisons are based on families which owned such cattle before and after the famine. Relevant figures, by occupational groups, are given in Table 6.8 which shows the proportion of families in the sample owning plough cattle before and after the famine in April 1943 and April 1944.

TABLE 6.8. NUMBER AND PROPORTION OF FAMILIES OWNING PLOUGH CATTLE IN APRIL 1943 AND APRIL 1944 IN EACH OCCUPATIONAL GROUP.

occupational groups	total in 1943	number of families surveyed			percentage of families			
		owning plough cattle		losing all plough cattle between April 1943 & April 1944	owning plough cattle		losing all plough cattle between April 1943 & April 1944 to families in April 1943	
		April 1943	April 1944		April 1943	April 1944	total families	owning families
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
agriculture	4852	3288	3077	211	69·0	64·2	4·8	6·4
agriculture & labour	2425	988	863	125	41·8	36·9	4·9	12·6
agricultural labour	2598	179	128	51	7·7	5·1	2·6	28·5
non-cultivating owner	886	252	228	24	28·4	25·5	2·9	9·5
others	4481	485	415	70	11·0	9·5	1·5	14·4

Col. (6) and (7) of the table show that both before and after the famine the proportions of families owning plough cattle was heaviest in the group 'agriculture' (69.0% and 64.2%) and next in 'agriculture & labour' (41.8% and 36.9%) before and after the famine. A fair proportion (28.4% before and 25.5% after the famine) of 'non-cultivating owner' had plough cattle but to a much less extent than the two other groups. In 'others' (11.0% and 9.5%) and 'agricultural labour' (7.7% and 5.1%) groups the proportion of families owning plough cattle was negligibly small.

6.14. The above distribution is the natural consequence of the manner in which the different occupational groups have been defined. The two groups 'agriculture' and 'agriculture & labour'depend mainly on the cultivation of their own land or land shared on *barga*; and they naturally must maintain plough cattle for this purpose. 'Non-cultivating owner', by definition, have their land cultivated on a share-cropping basis, and need not therefore maintain many plough cattle. 'Others' representing non-agricultural groups also naturally do not require plough cattle. Finally, 'agricultural labour' earn their living on a wage basis, and they have to do this because they can not afford to keep plough cattle. In fact if they did have plough cattle they would have immediately tried to obtain land on *barga* (share-croping) and thus moved up into the higher economic level of 'agriculture & labour'.

6.15. The seriousness of the position even before the famine can be appreciated from the fact that roughly one-fourth and one-half of the families in the groups 'agriculture' and 'agriculture & labour' respectively did not own any plough cattle. Even in normal times the cultivating groups themselves had to depend appreciably on hired cattle labour. Col.(8) shows that, proportionately to their total numbers, it is among these two groups that the largest number of families (4.8% and 4.9%) had lost all their plough cattle during the famine.

6.16. The relative deterioration was, however, different. This is shown in col.(9) of Table 6.8 which gives the percentage of families (in each occupational group) which owned plough cattle before the famine but had lost all such cattle during the famine. The loss was heaviest (28.5%) among 'agricultural labour', quite heavy among 'others' (14.4%) and 'agriculture & labour' (12.6%); and 9.5% among 'non-cultivating owner', and lowest in the 'agriculture' group (6.4%). On the whole, the loss was relatively highest in groups which owned less of plough cattle before the famine. Anomalies of distribution were thus further aggravated by the famine.

6.17. *Summary of conclusions*. The position of plough cattle in rural Bengal among agricultural groups is summarised below.

(1) There is a small over-all shortage which is aggravated by anomalies of distribution. In the three major groups ('agriculture', 'agriculture & labour', and 'agricultural labour') there is a surplus which would be wasted unless they get enough additional land on *barga*.

(2) The mean number of plough cattle owned per family decreased in all the groups, but the reduction was largest in 'agriculture' and 'agriculture & labour', the two groups mainly concerned with cultivation.

(3) Total loss of cattle was also heavist in the two major groups 'agriculture' and 'agriculture & labour'. Even before the famine a considerable number of these families did not own any cattle; their number increased during the famine which is likely to have serious consequences on agriculture in Bengal in future.

CHAPTER 7. ECONOMIC BACKGROUND OF THE FAMINE

7.1. A general review of the disintegration of rural economy of Bengal brought about by the famine in 1943 has been given in the foregoing chapters. The sample survey shows that 3.5 lakhs of families consisting of about 11 lakhs of persons had been reduced to destitution (para 3.5); about 7 lakhs of families or 38 lakhs of persons had suffered economic deterioration (para 4.5); inequalities in the distribution of paddy land in the province had been accentuated by heavy selling and mortgaging of land especially in the lower economic groups (para 5.25); and serious difficulties had been created by the loss of plough cattle mainly through sales (para 6.5).

7.2. *Over-all effect of the famine.* Table 7.1 gives in a consolidated form the extent of disintegration in regard to the three main elements, viz. destitution, sale paddy land, and loss of plough cattle, considered in the present report.

TABLE 7.1. EXTENT OF INCIDENCE OF FAMINE CONDITIONS IN RESPECT OF DIFERENT ELEMENTS WITH COMPARTIVE INDEX NUMBERS FOR DIFFERENT CLASSES OF SUBDIVISIONS

	percentage of			index number of			
Industries Department classification	families becoming destitute between Jan. 1943 & May 1944	families selling paddy land between April 1943 & April 1944	plough cattle sold between April 1943 & April 1944	desti-tution	sale of paddy land	sale of plough cattle	general average
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Class A	1·97	12·5	14·5	188	139	128	152
Class B	0·77	9·8	13·2	73	109	116	99
Class C	0·71	6·3	8·2	68	·70	73	70
rural Bengal	1·05	9·0	11·3	100	100	100	100

Quantitative estimates for the three elements are given in cols. (2), (3) and (4). These are expressed in the form of index numbers in corresponding cols. (5), (6) and (7). For example, from the bottom line in col. (2) it is seen that the increase in the number of destitute families between January 1943 and May 1944 was 1.05% on an average for rural Bengal as a whole; this has been adopted as the standard for comparison and accepted as 100 for calculating index numbers. The observed increase in the percentage of destitute families in Class (A) was 1.97%; dividing 1.97% by the adopted standard 1.05% and expressing as a percentage is obtained the index number of 188 given in col. (5). The other figures in cols. (5), (6) and (7) are obtained in the same way. Such index numbers supply a general over-all picture. It must be remembered, however, that their use is limited to certain special purposes and also that their sampling fluctuations are often not accurately known.

7.3. Cols. (2)—(4) and corresponding cols. (5)—(7) of Table 7.1 show that the extent of disintegration in respect of each element was greatest in Class (A) comprising the subdivisions which were most severely affected by famine conditions, less marked in Class (B) comprising the next group of subdivisions, and still less in Class (C) consisting of subdivisions which were believed to be only slightly affected.

7.4. Col. (8) gives the average index numbers for each of the three classes (A), (B) and (C), from which it is seen that for class (A) the generalized index number is highest and 152 ; for class (B) it is 99 which is about the same as the average for rural Bengal as a whole ; and finally for class (C) the index number is 70 which is distinctly less than the average. It is of some interest to note that destitution was relatively highest in class (A) with an index number of 188 but was not widely different in classes (B) and (C), which two classes were, however, appreciably differentiated in sales of paddy land and loss of plough cattle.

7.5. *Basis of the famine.* In respect of all elements of disintegration, the Industries Department classification was most in accord with the results of the survey. One thing is clear ; there were large differences in the intensity of incidence of the famine in different regions. Certain areas were very severely affected ; others to a moderate extent ; and still others only to a slight extent. This patchy nature of the incidence of famine conditions suggests that large regional differences in economic conditions had already existed before the famine, and were further accentuated during the crisis. This is corroborated by differences in the amount of paddy land owned, on an average, by each family (para 5.5) as well as by differences in rates of destitution (para 3.4) which had existed in different areas in January 1943.

7.5. *Differential incidence among occupational groups.* Table 7.1 shows differences in incidence in different geographical regions. There were also large differences in the effect of the famine in different occupational groups. Comparative statistics regarding destitution and economic deterioration have been given in Table 7.2. Index numbers of destitution shown in col. (5) of Table 7.2 are the same as those given in col. (4) of Table 3.6. Index numbers of economic deterioration have been constructed in the same way as index numbers of destitution (shown in Table 3.6) and are given in cols. (2)—(4) of Table 7.2.

Table 7.2. Index numbers of destitution and economic deterioration between January 1943 and May 1944 by occupational groups as in January 1943

occupational groups	proportion of families		index number		
	living in January 1943	suffering economic deterioration between January 1943 & May 1944	economic deterioration $\left(\frac{100\times \text{col.}(3)}{\text{col.}(2)}\right)$	index of destitution [col. (4) of table 3.6]	average
(1)	(2)	(3)	(4)	(5)	(6)
agriculture	32·46	27·88	85	22	54
agriculture & labour	16·85	21·90	131	24	78
agricultural labour	16·66	11·39	67	89	78
non-cultivating owner	6·11	1·75	28	30	29
others[1]	25·18	37·08	133	62	98
fishing	1·36	2·19	161	223	192
craft	4·97	4·96	99	77	88
husking paddy	1·72	0·44	25	79	54
transport	0·66	1·02	154	75	114
trade	6·73	23·07	372	37	204
profession & service	6·59	4·23	64	40	52
non-agricultural labour	0·98	1·17	119	51	85
other productive occupations	2·17	—	—	52	52
living on charity	2·74	—	—	—	—
total	100·00	100·00			

1) 'Others' excludes the group 'living on charity'

7.6 The index numbers reveal many features of interest from the point of view of rehabilitation. First consider the average values in col. (6). It is noted that rural 'trade' and 'fishing' groups were the worst sufferers with average index numbers of 204 and 192 respectively; rural 'transport' had also been severely affected with an average index number of 114. 'Craft' (88), 'non-agriculture labour' (85), 'agricultural labour' (78), and 'agriculture & labour' (77) were all seriously affected by destitution and economic deterioration. 'Agriculture' (53), 'husking paddy' (54), and "non-cultivating owner' (29) were on the other hand least affected by famine conditions.

7.7. Comparing the figures in cols. (4) and (5) it is interesting to note that economic deterioration was a more important factor than destitution in the groups 'agriculture' and 'agriculture & labour' while the reverse had been the case in regard to 'agricultural labour' and 'others'. Among 'others' again, economic deterioration was relatively more important in the groups of rural 'trade', 'transport', 'non-agricultural labour', 'profession & service and 'craft' while destitution was relatively more important in the groups 'husking paddy" and 'fishing'. Among 'non-cultivating owner' both factors were comparatively small and nearly equal. Occupational groups which could live on their assests for a time were evidently able to resist destitution to some extent, but groups in a more precarious position had succumbed more quickly to famine conditions.

7.8 *Differential incidence among agricultural groups.* Conditions in the agricultural groups are shown in greater detail in table 7.3. The index numbers regarding land alienation and mortgage [shown in cols (4)—(6) of the table] and those regarding loss of plough cattle [shown in cols (7) and (8)] are derived in the same way as index numbers of destitution (table 3.7) and of economic deterioation (table 7.2). Actual calculations for index numbers of land alienation and mortgage are shown in the table 7(A-1) and for index numbers of plough cattle in table 7(A-2).

TABLE 7.3. INDEX NUMBERS OF DESTITUTION, ECONOMIC DETERIORATION, FAMILIES AFFECTED BY SALE AND MORTGAGE OF PADDY LAND AND COMPLETE LOSS OF PLOUGH CATTLE, AND NET LOSS OF PLOUGH CATTLE DURING 1943 BY OCCUPATIONAL GROUPS AS IN JANUARY 1943

occupational groups as in January 1943	index number of disintegration in 1943							
	destitution	economic deterioration	families affected by			plough cattle		total
			sale of paddy land		mortgage of paddy land	net loss in number	families losing completely	
			in full	in part				
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
agriculture	22	85	56	122	99	83	69	77
agriculture & labour	24	131	163	95	120	131	137	11
agricultural labour	89	67	263	134	176	231	307	181
non-cultivating owner	30	28	44	51	56	97	103	58
others	62	133	13	69	88	135	156	111

From col. (9) of the table which gives the average values of the index numbers regardingall the four elements considered in cols. (2)—(8) (viz. destitution, economic deterioration, land alienation and mortgage, and loss of plough cattle) it is clear that on the whole 'agricultural labour' (with an average index number of 181) had suffered most severely from the effect of the famine ; 'agriculture & labour' came next with an index number of 114 ; agriculture' (index number 77), and 'non-cultivating owner' (index number 58) on the whole had suffered much less because of their superior status. In comparison 'others' (index number 111) came below 'agricultural labour' (181) but near the group 'agriculture & labour' (114).

7.9. Coming to individual elements, it is noticed from col. (4) of table 7.3 that sale of paddy land in full was proportionately highest among 'agricultural labour' (index number, 263) and then among 'agriculture & labour' (163) but was comparatively less important in the group 'agriculture' (56) and 'non-cultivating owner' (44). The 'others' (i.e. non-agricultüral groups) also were obliged to sell a great deal of their paddy land, and in fact, with an index number of 131, were affected in this respect much more seriously than 'agriculture' and 'non-cultivating owner'. Sale of paddy land in part (col. 5 of the table) was, however, relatively highest in 'agriculture (122) and 'agricultural labour' (134), and moderately high in 'agriculture & labour' (95) but low in 'others' (69) or 'non-cultivating owner' (51). Mortgaging of paddy land (col. 6) was most important in 'agricultural labour' (index number 176) ; quite high in 'agriculture & labour' (120), moderate in 'agriculture' (99) and 'others' (88), and least important in the group 'non-cultivating owner' (50). Loss of plough cattle, it is seen from col. (7) of the table, was highest in 'agricultural labour' (index number 231), quite severe in 'agriculture & labour' (131), and 'others' (135), and moderate in 'agriculture' (83) and 'non-cultivating owner' (97). Complete loss of plough cattle (for which index numbers are shown in col. 8) shows a more or less parallel incidence.

7.10. It is clear from the table that the group 'agricultural labour' (i.e. landless and unskilled labour) had suffered most severely in every respect with proportionately the highest number of sales and mortgaging of paddy land as well as of loss of plough cattle. 'Agriculture & labour' had high sales in full and also of mortgaging of paddy land as well as heavy loss of plough cattle. 'Agriculture' and 'non-cultivating owner', on the other hand, had comparatively more of sales in part of paddy land and less loss of plough cattle. 'Others' (i.e. non-agricultural group) had relatively high sales in full of paddy land and also comparatively heavy loss of plough cattle.

7.11. *Economic deterioration before the famine.* Changes in economic conditions in rural Bengal during the famine period of January 1943 to May 1944 have been already discussed in Chapters 3 and 4. Similar changes which had occurred during the so-called normal pre-famine period (from January 1939 to January 1943) are shown in table 7.4 in which the arrangement of columns is exactly the same as in table 4.5. Table 7.4 has been prepared from data given in table 7(A-3) which shows the estimated total number of families under each family occupation in January 1939 and in January 1943 in rural Bengal. Corresponding percentages are shown in table 7(A-4) and the sample surveyed in le atb7 (A-5).

TABLE 7.4. ESTIMATED NUMBER OF FAMILIES IN LAKHS IN RURAL BENGAL BY FAMILY OCCUPATION IN JANUARY 1939, AND CHANGE IN ECONOMIC CONDITION BETWEEN JANUARY 1939 AND JANUARY 1943.

occupational groups	estimated number of famlies (in lakhs)					relative change (expressed as percentage of families in January 1939) between January 1939 & January 1943			
	in January 1939	change between January 1939 & January 1943							
		improve-ment	deterio-ration	desti-tution	ambi-guous	improve-ment	deterio-ration	desti-tution	ambi-guous
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
agriculture	32·7	—	2·5	0·2	—	—	7·65	0·61	—
agriculture & labour	16·5	0·4	1·1	0·1	—	2·42	6·67	0·61	—
agriculture labour	17·5	1·2	0·6	0·3	—	6·86	3·43	1·71	—
non-cultivating owner	4·9	0·1	0·2	—	—	2·04	4·08	—	—
fishing	1·8	0·5	—	—	—	27·78	—	—	—
craft	5·3	—	0·4	0·2	0·4	—	7·55	3·77	7·55
husking paddy	1·2	—	0·1	0·1	—	—	8·33	8·33	—
transport	1·3	0·2	0·5	—	—	15·38	38·46	—	—
trade	8·2	—	0·6	0·1	1·1	—	7·32	1·22	13·41
profession & service	7·9	0·5	1·0	0·1	0·2	6·33	12·66	1·26	2·53
non-agricultural labour	1·0	0·2	—	—	0·1	20·00	—	—	10·0
others with productive occupations	2·3	0·2	—	—	0·2	8·70	—	—	8·70
living on charity	1·8	0·1	—	—	—	5·56	—	—	—
total	102·4	3·4	7·0	1·1	2·0	3·32	6·84	1·07	1·95

Cols. (7)—(10) of the table show that during the so-called normal period (January 1939 to January 1943) the proportion of families suffering economic deterioration and destitution was much higher than the proportion who had improved their position. From the bottom line it is seen that 6.84% of families had suffered economic deterioration against 3.32% who had improved their position, and 1.07% whose position was not clear. This shows conclusively that the economic deterioation had set in definitely in the pre-famine period; and that the famine itself was its culmination.

7.12. *Accelerated changes during the famine.* Table 7.5 shows relative changes in the pre-famine (January 1939—January 1943) and in the famine (January 1943—May 1944) periods. The proportion (in percentage form) of families in each category of 'improvement', 'no-change', 'deterioration', 'destitution', and 'not clear' in the pre-famine period (January 1939—January 1943) are shown in col.(2) and corresponding percentages of families for changes which occurred during the famine period (January 1943—May 1944) in col. (3). (It may be mentioned here that figures in col. (2) of this table have been obtained from the bottom line of table 7.4 and figures in col. (3) from the bottom line of table 4.5). Finally, figures in col. (3) have been expressed as percentages of corresponding figures in col. (2) in the form of index numbers which are shown in col. (4) as 'relative change'.

TABLE 7.5. ESTIMATED PROPORTION OF FAMILIES SHOWING SPECIFIED CHANGE IN ECONOMIC CONDITION BEFORE AND DURING FAMINE.

nature of change in economic condition of family	percentage of total families		relative change [col.(3) as percentage of col. (2)]	relative rate of change in the two periods
	January 1939—January 1943	January 1943—May 1944		
(1)	(2)	(3)	(4)	(5)
improvement	3·3	2·3	70	209
no change	86·8	85·9	99	—
deterioration	6·8	6·8	100	300
destitution	1·1	4·4	400	1200
not clear	2·0	0·6	30	—
total	100·0	100·0		

Col. (4) of the table shows interesting features. The relative proportion of families who had improved their economic position in the famine period was only 70% of those who had done so in the pre-famine period. The status quo was maintained by roughly 85% or 86% of all families in the two periods. The relative proportion of families suffering deterioration was the same in the two periods (relative change 100 in col. 4), but the relative proportion of families rendered destitute was four times greater during the famine period.

7.13. These are results based on the total period in the two epochs. The number of months covered in the pre-famine epoch was, however, 48 or 3 times greater than the number of months, namely, 16 covered by the famine period. Relative changes during the whole of the two periods shown in col.(4) must be, therefore, multiplied by 3 to give proportional rates of change in the same length of time. These figures are given in col. (5) of the table.

7.14. The greatly accelerated rates of change during the famine period are now quite clear. Relatively speaking, the time-rate of 'improvement' during the famine period was twice as great as that in the pre-famine period ; but this was set off by the three times greater rate of deterioration, and twelve times greater rate of destitution. The famine period was thus one of violent economic changes. 'Improvement' of economic conditions was also quicker but was restricted to a comparatively small number of families. 'Deterioration' and 'destitution' were even more violently accelerated and were shared by a much larger number of families. It is interesting to note, however, that roughly 85% or 86% of families succeeded broadly in maintaining their status quo. This shows that there was a great deal of economic inertia in rural Bengal which persisted even under famine conditions.

7.15. To put the matter very briefly, economic deterioration, on the whole, had set in definitely even in the pre-famine period (January 1939—January 1943) ; a comparatively small number of families were improving their economic position while a far large number were suffering economic deterioration or destitution. During the famine period (January 1943—May 1944) the whole process was greatly accelerated but the general nature of changes remained much the same. A small number of families improved their position while much larger number were impoverished or were rendered destitutes.

APPENDIX TO CHAPTER 7

TABLE 7.(A-1). RELATIVE PROPORTION OF FAMILIES SELLING AND MORTGAGING PADDY LAND BETWEEN APRIL 1943 AND APRIL 1944 TO THE PROPORTION OF FAMILIES OWNING PADDY LAND UNDER OCCUPATIONAL GROUPS AS IN JANUARY 1943.

occupational groups as in January 1943	percentage to total families				index of change		
	owning in April 1943	selling between April 1943 & April 1944		mortgaging between April 1943 & April 1944	sale		mortgage
		in full	in part		in full	in part	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
agriculture	46·34	26·03	56·63	45·67	56	122	99
agriculture & labour	20·76	33·84	19·66	24·84	163	95	120
agricultural labour	4·29	11·28	5·73	7·53	263	134	176
non-cultivating owner	9·85	4·34	5·06	5·52	44	51	56
others	18·76	24·51	12·92	16·44	131	69	88
total	100·00	100·00	100·00	100·00	—	—	—

TABLE 7.(A-2). RELATIVE PROPORTION OF NET LOSS OF PLOUGH CATTLE AND FAMILIES LOSING PLOUGH CATTLE COMPLETELY BETWEEN APRIL 1943 AND APRIL 1944 TO THE PROPORTION OF TOTAL PLOUGH CATTLE AND FAMILIES OWNING PLOUGH CATTLE UNDER OCCUPATIONAL GROUPS AS IN JANUARY 1943.

occupational groups as in January 1943	number of plough cattle		percentage to total number of plough cattle			percentage to total families		
	owned in April 1943	net loss between April 1943 & April 1944	owned in April 1943	net loss between April 1943 & April 1944	index of change	owning in April 1943	losing in total between April 1943 & April 1944	index of change
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
agriculture	7614	869	66·73	55·14	83	63·33	43·87	69
agriculture & labour	1875	338	16·43	21·45	131	19·03	25·99	137
agricultural labour	332	106	2·91	6·73	231	3·45	10·60	307
non-cultivating owner	627	84	5·50	5·33	97	4·85	4·99	103
others	962	179	8·43	11·35	135	9·34	14·55	156
total	11410	1576	100·00	100·00		100·00	100·00	—

Table 7.(A-3) Estimated number of families (in lakhs) under different occupations in rural Bengal in January 1939 and in January 1943

family occupations in January, 1939.	estimated number of families (in lakhs) under respective occupations in January 1943													
	agri-culture	agri-culture & labour	agri-cultural labour	non-culti-vating owner	fishing	craft	husking paddy	trans-port	trade	profe-ssion & service	non-agri-cultural labour	others with produc-tive occupa-tions	living on charity	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
agriculture	30·0	1·1	0·6	0·3	—	—	0·1	0·1	0·1	0·1	0·1	—	0·2	32·7
agriculture & labour	0·4	14·9	0·6	0·1	—	0·2	—	—	—	0·1	—	0·1	0·1	16·5
agricultural labour	0·4	0·6	15·5	0·1	—	0·1	0·1	—	0·1	0·2	—	0·1	0·3	17·5
non-cultivating owner	0·1	—	0·1	4·6	—	—	—	—	—	0·1	—	—	—	4·9
fishing	0·1	0·1	—	—	1·3	0·3	—	—	—	—	—	—	—	1·8
craft	0·4	—	0·1	0·1	—	4·3	—	—	—	0·1	—	0·1	0·2	5·3
husking paddy	0·1	—	—	—	—	—	1·0	—	—	—	—	—	0·1	1·2
transport	0·1	0·1	—	0·1	—	—	0·4	0·6	—	—	—	—	—	1·3
trade	1·0	0·1	0·1	0·2	—	0·1	—	—	6·4	0·1	0·1	—	0·1	8·2
profession & service	0·5	0·1	0·2	0·5	—	0·1	0·1	—	0·1	6·1	0·1	—	0·1	7·9
non-agricultural labour	0·1	—	—	0·1	—	—	—	—	0·1	—	0·7	—	—	1·0
others with productive occupations	0·1	0·1	0·1	0·1	—	—	—	—	—	—	—	1·9	—	2·3
living on charity	—	—	—	—	—	—	—	—	0·1	—	—	—	1·7	1·8
total	33·3	17·1	17·3	6·2	1·3	5·1	1·7	0·7	6·9	6·8	1·	2·2	2·8	102·4

TABLE 7.(A-4). WEIGHTED PERCENTAGE OF FAMILIES UNDER DIFFERENT OCCUPATIONS IN JANUARY 1939 AND IN JANUARY 1943.

occupational groups as in January 1939	weighted percentage of families under respective occupations in January 1943													
	agri-culture	agri-culture & labour	agri-cultural labour	non-cultiva-ting owner	fishing	craft	husking paddy	trans-port	trade	profes-sion & service	non-agri-cultural labour	others with productive occupa-tions	living on charity	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
agriculture	29·26	1·04	0·56	0·28	—	0·04	0·04	0·03	0·13	0·09	0·04	0·03	0·18	31·72
agriculture & labour	0·35	14·56	0·61	0·06	0·01	0·15	0·02	0·01	0·03	0·08	0·01	0·05	0·11	16·05
agricultural labour	0·37	0·60	15·11	0·15	—	0·08	0·11	0·01	0·04	0·16	0·04	0·11	0·29	17·07
non-cultivating owner	0·15	0·03	0·04	4·51	—	0·01	0·03	—	0·01	0·05	0·01	0·03	0·03	4·90
fishing	0·07	0·05	0·03	0·01	1·35	0·32	—	—	0·01	—	0·01	0·01	0·03	1·89
craft	0·43	0·03	0·06	0·14	—	4·20	—	—	0·03	0·05	0·01	0·04	0·14	5·13
husking paddy	0·09	0·01	0·03	0·04	—	0·02	0·99	0·01	0·01	·01	0·02	0·01	0·04	1·28
transport	0·07	0·04	0·03	0·08	—	—	0·42	0·60	0·02	—	0·01	0·01	0·03	1·31
trade	1·01	0·14	0·14	0·22	—	0·05	0·03	—	6·27	0·13	0·05	0·02	0·07	8·13
profession & service	0·53	0·12	0·17	0·47	—	0·07	0·04	—	0·10	5·99	0·05	0·03	0·09	7·66
non-agricultural labour	0·07	0·01	—	0·05	—	0·01	—	—	0·01	0·01	0·70	—	0·01	0·87
others with productive occupations	0·07	0·04	0·04	0·10	—	0·01	0·01	—	0·03	0·01	0·01	1·81	0·03	2·16
living on charity	0·01	—	0·04	—	—	0·01	0·03	—	0·06	0·03	0·02	0·02	1·61	1·83
total	32·48	16·67	16·86	6·11	1·36	4·97	1·72	0·66	6·75	6·61	0·98	2·17	2·66	100·00

TABLE 7.(A-5). NUMBER OF FAMILIES SURVEYED UNDER DIFFERENT OCCUPATIONS IN JANUARY 1939 AND IN JANUARY 1943.

occupational groups as in January 1939	number of families surveyed under respective occupations in January 1943													
	agriculture	agriculture & labour	agricultural labour	non-cultivating owner	fishing	craft	husking paddy	transport	trade	profession & service	non-agricultural labour	others with productive occupations	living on charity	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
agriculture	4368	156	91	43	—	8	8	6	20	16	5	4	31	4756
agriculture & labour	41	2164	93	9	2	30	4	1	6	12	3	10	22	2397
agricultural labour	66	111	2252	25	—	12	20	1	7	25	6	15	45	2585
non-cultivating owner	25	5	8	619	—	1	4	—	3	9	2	3	5	684
fishing	12	9	4	1	218	67	—	—	1	—	1	1	5	319
craft	65	6	13	22	—	656	1	1	5	8	2	6	24	809
husking paddy	17	2	5	8	—	3	143	1	1	2	3	2	7	194
transport	11	6	5	10	—	—	58	105	4	—	1	1	5	206
trade	155	24	25	32	—	9	5	1	1009	21	9	4	14	1308
profession & service	91	20	27	72	—	11	6	—	19	962	9	6	17	1240
non-agricultural labour	11	2	—	5	—	1	—	—	1	1	88	—	2	111
others with productive occupations	12	6	5	14	—	1	1	—	3	3	1	272	6	324
living on charity	1	—	6	—	—	1	6	—	8	5	4	3	241	275
total	4875	2511	2534	860	220	800	256	116	1087	1064	134	327	424	15208

Paper received : 20 April 1946.

ON A MEASURE OF DIVERGENCE BETWEEN TWO MULTINOMIAL POPULATIONS.

By A. BHATTACHARYYA
Statistical Laboratory, Calcutta.

INTRODUCTION

The problem of discrimination in Statistics is not new. It has repeatedly been stressed by Mahalanobis (1930) that mere "tests of significance" can not always answer all our questions and for this reason he introduced the idea of divergence of two populations. His "generalised distance" measures the divergence of two normal correlated multivariate populations with the same sets of variances and covariances. Bose and Roy (1938) found the distribution of D², "Mahalanobis's generalised distance" in its studentised form.

2. DEFINITION OF DIVERGENCE

Let two multinomial populations be characterised by the two sets of probabilities $(\pi_1, \ldots \pi_k)$ and $(\pi'_1, \ldots \pi'_k)$. Then, as $\Sigma\pi = 1$ and $\Sigma\pi' = 1$, $(\sqrt{\pi_1}, \ldots \sqrt{\pi_k})$ and $(\sqrt{\pi'_1}, \ldots \sqrt{\pi'_k})$ may be considered to be the direction cosines of two st. lines through the origin in a k-dimensional space. The square of the angle between these two lines may be considered to be an appropriate measure of divergence between the two multinomial populations. Thus, if the measure of divergence be denoted by Δ^2 then

$$\cos\Delta = \sqrt{\pi_1\pi_1'} + \sqrt{\pi_2\pi_2'} + \quad . \quad . \quad . \quad + \sqrt{\pi_k\pi_k'} \qquad .. \quad (2.1)$$

This can be written as

$$4\sin^2\frac{\Delta}{2} = (\sqrt{\pi_1} - \sqrt{\pi_1'})^2 + (\sqrt{\pi_2} - \sqrt{\pi_2'})^2 + \quad . \quad . \quad . \quad + (\sqrt{\pi_k} - \sqrt{\pi_k'})^2 \qquad .. \quad (2.2)$$

From this latter expression it is evident that when $\pi_i = \pi_i'$ $(i = 1, 2, \ldots k)$ Δ vanishes. Also when Δ vanishes, we have conversely $\pi_i = \pi'_i$ $(i = 1, 2, \ldots k)$ that is, the two populations are identical.

3. DIVERGENCE BETWEEN A SAMPLE AND THE POPULATION, ITS RELATION TO PEARSONIAN χ^2 FOR LARGE n

Let n be the size of a sample and (p_i) be the observed proportions taken from a population with proportions (π_i). Then, from the definition of divergence, the divergence β^2 of the sample from the population is given by

$$\cos\beta = \sum_{i=1}^{k}\sqrt{p_i\pi_i} = \sum_{i=1}^{k}\pi_i\left[1 + \frac{p_i - \pi_i}{\pi_i}\right]^{\frac{1}{2}} \qquad .. \quad (3.1)$$

$$= \sum_{i=1}^{k}\pi_i\left[1 + \frac{p_i - \pi_i}{2\pi_i} - \tfrac{1}{8}\frac{(p_i - \pi_i)^2}{\pi_i^2} + \varepsilon_i\right]$$

where $|\varepsilon_i|$ is of the order of $\frac{1}{n^{3/2}}$

$$= \sum_{i=1}^{k}\pi_i + \tfrac{1}{2}\sum_{i=1}^{k}(p_i - \pi_i) - \tfrac{1}{8}\sum_{i=1}^{k}\frac{(p_i - \pi_i)^2}{\pi_i} + \sum_{i=1}^{k}\varepsilon_i\pi_i \qquad .. \quad (3.2)$$

$$= 1 - \tfrac{1}{8}\frac{\chi^2}{n} + \varepsilon$$

where $\chi^2=\Sigma(np_i-n\pi_i)^2/n\pi_i$ *the Pearsonian* χ^2 *and* ε *is* $O(1/n^{3/2})$. Thus

$$2\sin^2\frac{\beta}{2}=\tfrac{1}{8}\,\frac{\chi^2}{n} \qquad .. \quad (3.3)$$

approximately, when n is very large. Now we know that, for large values of k, $(\sqrt{2\chi^2}-\sqrt{2k-3})$ is distributed normally with unit standard deviation. So, χ^2 is of the order of k and χ^2/n is very small.

So, finally we get

$$\beta^2=\tfrac{1}{4}\,\frac{\chi^2}{n}\ \ .i.e,\ \beta=\tfrac{1}{2}\frac{\chi}{\sqrt{n}} \qquad .. \quad (3.4)$$

4. ASYMPTOTIC FORM OF THE MULTINOMIAL PROBABILITY WHEN n IS LARGE

We know that the multinomial probability

$$\frac{n\,!}{(np_1)\,!\,(np_2)\,!\ldots\ldots(np_k)!}\,\pi_1^{\,np_1}\,.\,\pi_2^{\,np_2}\ldots\ldots\ldots\ldots\pi_k^{\,np_k}$$

$$\longrightarrow \text{Const.}\ e^{-\chi^2/2}\,.\frac{dp_1}{\sqrt{\pi_1}}.\frac{dp_2}{\sqrt{\pi_2}}\ldots\ldots\frac{dp_k}{\sqrt{\pi_k}}$$

$$\longrightarrow \text{Const.}\,e^{-2n\beta^2}\,d\sqrt{p_1}.\,d\sqrt{p_2}\ldots.d\sqrt{p_k}$$

for $\sqrt{p_i}=\sqrt{\pi_i}+(p_i-\pi_i)\left[\dfrac{d\sqrt{\pi_i}}{d\pi_i}\right]+\dfrac{(p_i-\pi_i)^2}{2!}\left[\dfrac{d^2\sqrt{\pi_i}}{d\pi_i^{\,2}}\right]+\ldots\ldots\ldots$

(by Taylor's theorem) and so $d\sqrt{p_i}=dp_i/\sqrt{\pi_i}$ approximately, for $\pi_i\neq o$ and $(p_i-\pi_i)$ is $O(1/n^{3/2})$

$$\longrightarrow \text{Const.}\ e^{-2n\beta^2}.\,ds \qquad .. \quad (4.1)$$

where ds is an elementary volume given by $d\sqrt{p_1}\,d\sqrt{p_2}\ldots d\sqrt{p_k}$ subject to $\Sigma p=1$.

The above could be seen to be true also from the following considerations. We know that the multinomial distribution may be considered to be composed of a number of Poisson distributions.

Thus

$$\frac{n\,!}{(np_1)\,!\,(np_2)\,!\ldots\ldots(np_k)\,!}\pi_1^{\,np_1}\,.\,\pi_2^{\,np_2}\ldots\ldots\pi_k^{\,np_k}$$

$$=\frac{n\,!}{e^{-n}\,.\,n^n}\cdot\frac{e^{-n\pi_1}\,(n\pi_1)^{np_1}}{(np_1)!}\cdot\frac{e^{-n\pi_2}\,(n\pi_2)^{np_2}}{(np_2)!}\ldots\ldots\frac{e^{-n\pi_k}\,(n\pi_k)^{np_k}}{(np_k)\,!}$$

We also know (Bartlet : 1936) that in Poisson's distribution if x is the observed and a the expected value, then $\sqrt{x}$ is distributed about $\sqrt{a}$ normally with standard deviation 1/2. The above expression becomes

$$\text{Const.}\ e^{-2n(\sqrt{p_1}-\sqrt{\pi_1})^2}\,d\sqrt{p_1}\,.\,e^{-2n(\sqrt{p_2}-\sqrt{\pi_2})^2}\,d\sqrt{p_2}\ldots.e^{-2n(\sqrt{p_k}-\sqrt{\pi_k})^2}\,d\sqrt{p_k}$$

$$=\ \text{Const}\ e^{-2n\sum\limits_{i=1}^{k}(\sqrt{p_i}-\sqrt{\pi_i})^2}\cdot d\sqrt{p_1}\,.\,d\sqrt{p_2}\ldots\ldots\ldots\ldots d\sqrt{p_k}$$

$$\sim\ \text{Const.}\ e^{-2n.\,4\sin^2\frac{\beta}{2}}\,ds$$

$$\sim\ \text{Const.}\ e^{-2n\beta^2}\,ds$$

he same as (4.1), on the supposition that β is very small. .. (4.2)

5. GEOMETRICAL INTERPRETATION OF THE VOLUME ELEMENT

As the point $(\sqrt{p_1} \dots \sqrt{p_k})$ lies on the unit hypersphere in the k-dimensional space, $d\sqrt{p_1} \dots d\sqrt{p_k}$ is an elementary surface volume of a sphere round the sample point, say S. With this geometrical representation, the probability of a sample is completely representable on the unit hypersphere. Assuming the law to hold for all values of β, we may find the distribution of β. For keeping β fixed if we integrate out over the available space on the sphere, we get something like a zone of annular space. The volume of this space is proportional to $\beta^{k-2}d\beta$ (for the volume of the zone is proportional to $\sin^{k-2}\beta$ and so approximately to $\beta^{k-2}d\beta$). So, the distribution of β is

$$\text{Const.}\ e^{-2n\beta^2}\ \beta^{k-2}\, d\beta \qquad \dots \quad (5.1)$$

which is also obvious from the relation between β^2 and χ^2.

6. DEFINITION OF DIVERGENCE BETWEEN TWO SAMPLES

Corresponding to Δ^2 in the populations, we have the statistic defined by

$$\cos D' = \sqrt{p_1p_1'} + \sqrt{p_2p_2'} + \dots\dots\dots\dots + \sqrt{p_kp_k'}$$

where n, n' are the sizes and (p_i) and (p'_i) are the sample proportion. This is not an unbiassed estimate of Δ^2.

7. RELATION OF D'^2 WITH PEARSONIAN χ^2 WHEN $\Delta = o$ AND n AND n' ARE LARGE AND EQUAL TO ONE ANOTHER

When the sample sizes are the same, namely n, we can express D'^2 in terms of Pearsonian χ^2. For

$$\cos D' = \sum_{i=1}^{k} \sqrt{p_ip_i'} = \tfrac{1}{2}\sum_{i=1}^{k}\sqrt{(p_i+p_i')^2-(p_i-p_i')^2}$$

$$= \tfrac{1}{2}\sum_{i=1}^{k}(p_i+p_i')\left\{1-\frac{(p_i-p_i')^2}{(p_i+p_i')^2}\right\}^{\frac{1}{2}}$$

$$= \tfrac{1}{2}\sum_{i=1}^{k}(p_i+p_i')\left\{1-\tfrac{1}{2}\frac{(p_i-p_i')^2}{(p_i+p_i')^2}-\tfrac{1}{8}\cdot\frac{(p_i-p_i')^4}{(p_i+p_i')^4}-\dots\right\}$$

by the binomial theorem, because $\dfrac{(p_i-p_i')}{(p_i+p_i')} < 1$,

$$= \tfrac{1}{2}\sum_{i=1}^{k}(p_i+p_i')-\tfrac{1}{4}\sum_{i=1}^{k}\frac{(p_i-p_i')^2}{(p_i+p_i')}-\tfrac{1}{16}\Sigma\varepsilon_i$$

$$= 1-\tfrac{1}{4}\sum_{i=1}^{k}\frac{(p_i-p_i')^2}{(p_i+p_i')}-\varepsilon \qquad \dots \quad (7.1)$$

Now, Δ^2 being zero, $(p_i-p'_i)$ is of the order of $1/\sqrt{n}$ and so ε is of the order of $1/n^2$. So, when n is very large,

$$2\sin^2\frac{D'}{2} = \tfrac{1}{4}\sum_{i=1}^{k}\frac{(p_i-p_i')^2}{p_i+p_i'} \qquad \dots \quad (7.2)$$

The sample values are np_i and np'_i $(i=1, 2, \dots k)$. So from the usual method of calculating χ^2 we get

$$\chi^2=\sum_{i=1}^{k}\frac{\left(np_i-\frac{np_i+np_i'}{2}\right)^2}{\frac{np_i+np_i'}{2}}+\sum_{i=1}^{k}\frac{\left(np_i'-\frac{np_i+np_i'}{2}\right)^2}{\frac{np_i+np_i'}{2}}$$

$$=\sum_{j=1}^{k}\frac{(np_i-np_i')^2}{np_i+np_i'} \qquad \dots (7.3)$$

This χ^2, we know, is of the order of $2k$. So, χ^2/n is small and as from (7.2) $2sin^2\ D'/2=\chi^2/4n$ we get $D'^2=\chi^2/2n$. From this relation it is evident that on the null hypothesis, the distribution of D'^2 follows the distribution of χ^2 for very large samples.

8. DISTRIBUTION OF D'^2 WHEN $\Delta^2 \neq 0$

Let there be two samples with sizes n and n' and proportions (p_i) and (p'_i) taken respectively from the two populations characterised by (π_i) and (π'_i). Let β^2 and β'^2 be the divergences (i.e. square of the angular distances) of the two samples from their respective populations. Then the joint distribution of the two samples is

$$\text{Const.}\ e^{-2\beta n^2} ds\ .\ e^{-2n'\beta'^2} ds' \qquad \dots (8.1)$$

where ds and ds' are elementary surface volumes, on the unit hypersphere, about the sample points. Let P and P′ be the population points (i,e, where the population lines meet the unit sphere), S and S′ the sample points (with the same convention as the population points). Then Δ is represented by P P′, D′ by SS′, and β and β' by S P and S′P′. From the distributions of β and β' we know, when n and n' are very large, that these tend to become very small. That is, the probability of a large β is so small that, for all practical purposes, β may be regarded to lie within a narrow zone about the arc PP′. Similar is the case for β'. Outside this zone the probability of getting samples is negligible. This narrow zone may be conceived to be a portion of a cylinder (that is, the projection of this zone on the enveloping cylinder osculating the unit sphere along the curve PP′ may be regarded to be coincident with the zone itself). We may develope out this cylindrical zone into a flat of $(k-1)$ dimensions where the geodesics β, β', Δ and D′ all become right lines and ds and ds' become elementary areas (i,e. volumes) in this flat of $(k-1)$ dimensions.

With the help of the above geometrical device we may proceed to find the distribution of D′ which now represents a rectilinear length. In the $(k-1)$ flat space let the coordinates of the points P, P′, S and S′ be (a_i), (a'_i), (x_i) and (x'_i), respectively $(i=1, 2, \dots k-1)$. The joint probability of the two samples is

$$\text{const.}\ e^{-2n\beta^2-2n'\beta'^2}\ ds\ ds'$$

$$=\ \text{const.}\ e^{-2n\sum_{i=1}^{k-1}(x_i-a_i)^2\ -2n'\sum_{i=1}^{k-1}(x_i'-a_i')^2}\ \prod_{i=1}^{k-1}dx_i\ \prod_{i=1}^{k-1}dx_i'$$

From this it is easy to deduce the distribution of $(x_i - x'_i - a_i + a'_i)$ in the form

$$\text{Const.}\ e^{-\frac{2n\,n'}{n+n'}\left\{\sum_{i=1}^{k-1}(x_i - x_i' - a_i - a_i')^2\right\}} \prod_{i=1}^{k-1} d\,(x_i - x_i')$$

Or, which is the same as

$$\text{Const.}\ e^{-\frac{2nn'}{n+n'}\left\{\sum_{i=1}^{k-1}(x_i - x'_i)^2 - 2\sum_{i=1}^{k-1}(x_i - x_i')(a_i - a_i') + \sum_{i=1}^{k-1}(a_i - a_i')^2\right\}} \prod_{i=1}^{k-1}(x_i - x_i')$$

From this by adopting the method employed by Mr. R. C. Bose for getting the distribution of the classical D^2 (Bose 1936) we get from the above

$$\text{Const.}\ e^{-\frac{2nn'}{n+n'}\left\{D'^2 - 2D'\Delta\cos\theta + \Delta^2\right\}} (D'\sin\theta)^{k-3} D' dD' d\theta$$

Where $D'^2 = \sum_{i=1}^{k-1}(x_i - x_i')^2$, $\Delta^2 = \sum_{i=1}^{k-1}(a_i - a'_i)^2$ and $D'\Delta\cos\theta = \sum_{i=1}^{k-1}(x_i - x_i')(a_i - a_i')$

Integrating over θ from 0 to 2π we get the distribution of D' as

$$\text{Const.}\ D'^{(k-1)/2} e^{-\frac{2nn'}{n+n'}(D'^2 + \Delta^2)}\ I_{\frac{k-3}{3}}\left(\frac{4nn'}{n+n'} D'\Delta\right)$$

where I is a Bessel function of purely imaginary argument.

9. APPROXIMATION OF Δ AND D' WHEN Δ IS VERY SMALL

We know

$$\cos\Delta = \Sigma\sqrt{\pi_i \pi'_i} = \tfrac{1}{2}\,\Sigma(\pi_i + \pi'_i)\left[1 - \left(\frac{\pi_i - \pi'_i}{\pi_i + \pi_i'}\right)^2\right]^{\frac{1}{2}}$$

Replacing $\cos\Delta$ by $1 - \Delta^2/2$ to a first approximation we have $\Delta^2 = \Sigma(\pi_i - \pi_i')^2/(\pi_i + \pi'_i)$. Similarly $D'^2 = \Sigma(p_i - p'_i)^2/(p_i + p'_i)$. These were suggested by Prof. Mahalanobis as a measure of divergence between two multinomial populations and samples therefrom.

10. SPECIAL CASE OF BINOMIAL POPULATIONS

Let there be two binomial populations defined by (π_1, π_2) and (π_1', π'_2) from which two samples of sizes n and n' with proportions (p_1, p_2) and (p'_1, p'_2) are taken respectively. Then we know that $n(p_1 - \pi_1)\sqrt{n\pi_1(1-\pi_1)}$ is distributed normally about zero with unit standard deviation subject to the condition that n is sufficiently large and both π_1 and π_2 not far different from 1/2. Now, under the same assumptions, we have $(\sin^{-1}\sqrt{p_1} - \sin^{-1}\sqrt{\pi_1}) = (p_1 - \pi_1)/2\sqrt{\pi_1(\ -\pi_1)}$ approximately (by expanding in Taylor's series). Therefore $2\sqrt{n}\,(\sin^{-1}\sqrt{p_1} - \sin^{-1}\sqrt{\pi_1})$ is distributed normally about zero with unit standard deviation. Similarly, under the same set of conditions, $2\sqrt{n'}(\sin^{-1}\sqrt{p'_1} - \sin^{-1}\sqrt{\pi'_1})$ is distributed normally with unit standard deviation. So, from these two it easily follows that

$$2\sqrt{\frac{n\,n'}{n+n'}}\left\{\left(\sin^{-1}\sqrt{p_1} - \sin^{-1}\sqrt{p_1'}\right) - \left(\sin^{-1}\sqrt{\pi_1} - \sin^{-1}\sqrt{\pi_1'}\right)\right\}$$

is distributed with unit standard deviation. Here the population divergence is clearly $\sin^{-1}\sqrt{\pi_1}-\sin^{-1}\sqrt{\pi'_1}$ and the sample estimate is $\sin^{-1}\sqrt{p_1}-\sin^{-1}\sqrt{p'_1}$. Clearly this is a particular case of the general problem which degenerates into it when $k=2$.

References

1. Bartlett, M. S. (1936). "The Square-root Transformation in the Analysis of Variance". *J. R. S. S. (Supplement)*, 3, Pp. 68-78.
2. Bose, R. C. On tne exact distribution and moment co-efficients of the D^2 statistic. *Sankhya* 1(2) Pp. 143-54.
3. Bose, R. C. and Roy, S. N. (1938). "The Distribution of the Studentised D^2-Statistics", *Sankhya*, 4(1). Pp. 19-38.
4. Cochran, W. G. (1940). "Analysis of variance when experimental errors follow the Poisson or Binomial Laws". *Annals of Math. Stat.* 11(3). Pp. 335-347.
5. Mahalanobis, P. C. (1930). "On the tests and measures of group divergences". *Jour. and Proc. of the Asiatic Society of Bengal.* (New series), 26. No. 4. Pp. 541-588.
6. ——— "On the generalised distance in statistics". *Proc. of the Nat. Inst. of Sciences of India*, 2(1), 1936, Pp. 49-55.

Paper received 3 July 1941

TESTS WITH DISCRIMINANT FUNCTIONS IN MULTIVARIATE ANALYSIS

By C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta.

INTRODUCTION

In a paper (Rao : 1946), the author has considered a general method of arriving at suitable studentised statistics which are to be used in tests of linear hypotheses when the observed set of variables, whose expectations are linear functions of unknown parameters, are independent and have the same variance. The principle is to take a linear compound, subject to some restrictions of parametric functions and maximise the ratio of its estimate to the standard error of this estimate.

These methods can be extended to the case of observations from multivariate correlated populations and it is found that tests of significance can, in many cases, be carried out with the use of published tables of *t* and F alone. The general problem of distribution connected with the statistics arrived at by the above principle has been discussed below and solutions to a few problems which appear to be new have been given.

2. THE PROBLEM OF DISTRIBUTION

In the problems considered by Fisher (1938, 1940), Bartlet (1939), the tests of significance concerning discriminant functions were derived by drawing an analogy with the general regression problem involving pseudovariates. In cases where the introduction of pseudovariates is not possible we may use a standard distribution derived below. Let

$$\begin{array}{cccc} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ x_{1s} & x_{2s} & \cdots & x_{ks} \end{array} \quad .. \quad (2.1)$$

be *s* sets of observations on *k* variates $x_1, \ldots x_k$ characterised by the probability differential

$$\text{Const. } e^{-\frac{1}{2}[(x_1-m)^2+x^2_2+\ldots+x^2_k]}dx_1dx_2 \ldots dx_k \quad .. \quad (2.2)$$

We may represent the *s* observations on the *i*-th variate $x_{i1}, \ldots\ldots x_{is}$ by a point P_i in an Euclidean space of *s* dimensions or by the vector $\overrightarrow{OP_i}$ where O is the origin. The whole sample of observations (2.1) may then be represented by vectors $\overrightarrow{OP_1}, \ldots\ldots \overrightarrow{OP_k}$ in this space. Let $\overrightarrow{OA}$ represent a vector of unit length along a line which makes equal angles with the coordinate axes. The vector $\overrightarrow{OP_i}$ consists of two components one along $\overrightarrow{OA}$ and the other orthogonal to $\overrightarrow{OA}$ so that products of vectors $\overrightarrow{OP_i}$ and $\overrightarrow{OP_j}$ contain contributions due to components in these two directions which may be represented by y_{ij} and b_{ij} respectively. In this case

$$\left.\begin{array}{l} y_{ij}=s\bar{x}_i\bar{x}_j \\ b_{ij}=\Sigma(x_{ir}-\bar{x}_i)(x_{jr}-\bar{x}_j) \end{array}\right\} \quad .. \quad (2.3)$$

The configuration of the observations (2.1) is diagramatically represented below

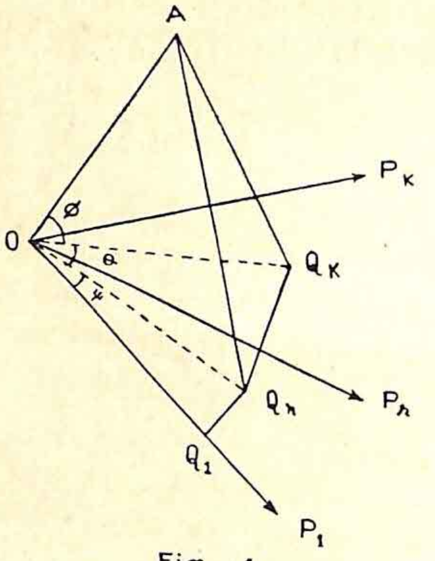


Fig 1.

where Q_i represents the foot of the perpendicular from A to the subspace constituted by the vectors $\overrightarrow{OP}_1, \ldots\ldots \overrightarrow{OP}_i$. The angles $A\hat{O}Q_k$, $Q_k\hat{O}Q_r$ and $Q_r\hat{O}Q_1$ are represented by ϕ, θ and ψ respectively. The magnitude of AQ_i is given by

$$AQ_i^2 = \frac{| b_{pq} |}{| b_{pq} + s\bar{x}_p\bar{x}_q |} \qquad (p,q = 1, 2, \ldots i) \qquad \ldots \quad (2.4)$$
$$= \frac{1}{1+V_i}$$

where $V_i = \sum_1^i \sum_1^i \underset{i}{b}^{pq} s\bar{x}_p\bar{x}_q$ and $\underset{i}{b}^{pq}$ are the elements of matrix $(\underset{i}{b}^{pq})$ reciprocal to (b_{pq}), $p, q = 1, 2, \ldots\ldots i$.

We are interested in the distributions of statistics V_r and $(1+V_k)/(1+V_r)$ given by

$$\left.\begin{aligned} V_r &= \frac{\cos^2\phi\cos^2\theta}{1-\cos^2\phi\cos^2\theta} \\ \text{and} \quad \frac{1+V_K}{1+V_r} &= 1+\cot^2\phi\sin^2\theta \end{aligned}\right\} \qquad \ldots \quad (2.5)$$

We may find the distributions of V_r and $U = [(1+V_k)/(1+V_r)]-1$. The joint distribution of ϕ, θ, ψ and t the length of $\overrightarrow{OP}_1$ is derivable by taking the product of the volumes contributed by allowing infinitesimal increments to ϕ, θ, ψ and t as

$$\begin{aligned} &\text{Const. } e^{-\frac{1}{2}t^2 + \sqrt{s}\, m\, t \cos\phi \cos\theta \cos\psi} \\ &\qquad t^{s-1}\, dt \\ &(\sin\phi)^{s-k-1} d\phi\, (\cos\phi \sin\theta)^{k-r-1} \cos\phi\, d\theta \\ &\qquad (\cos\phi\cos\theta\sin\psi)^{r-2} \cos\phi\cos\theta\, d\psi \end{aligned} \qquad \ldots \quad (2.6)$$

Integrating over ψ the above reduces to

$$\begin{aligned} &\text{Const. } e^{-\frac{1}{2}t^2}\, t^{s-1-(r-2)/2} \\ &\qquad I_{\frac{r-2}{2}}(2\sqrt{s}\, mt \cos\phi\cos\theta) \\ &(\sin\phi)^{s-k-1} (\cos\phi)^{k-1} (\cos\theta)^{r-1} (\sin\theta)^{k-r-1}\, dt\, d\phi\, d\theta \end{aligned} \qquad \ldots \quad (2.7)$$

Intergrating over t which varies from 0 to ∞ we get the distribution of ϕ and θ as

$$\text{Const.}\ (\sin\phi)^{s-k-1}(\cos\phi)^{k-1}(\sin\theta)^{k-r-1}(\cos\theta)^{r-1}$$
$$ {}_1F_1(s/2,\ r/2,\ 2sm^2\cos^2\phi\ \cos^2\theta)\ d\phi\ d\theta \qquad \ldots (2.8)$$

Changing over to the variables V_r and U we get their joint distribution as

$$\text{Const.}\ \frac{V_r^{\frac{r}{2}-1}}{(1+V_r)^{s/2-1}}\ {}_1F_1\left(s/2,\ r/2,\ 2\ sm^2\ \frac{V_r}{1+V_r}\right)dV_r \qquad \ldots (2.9)$$

$$\times\ \frac{U^{\frac{k-r}{2}-1}}{(1+U)^{\frac{s-r}{2}}}\ dU \qquad \ldots (2.10)$$

which shows that V_r and U are independently distributed. Since the distribution of V_r and U are directly derivable from the joint distribution of means, variances and covariances given by

$$\text{Const.}\,e^{-\frac{1}{2}\sum_{i}(s\bar{x}_i^2+b_{ii})+s\bar{x}_1m}\ |\,b_{ij}\,|^{\frac{s-1}{2}-\frac{k+1}{2}}\ \pi\ dx_i\ \pi\ db_{ij} \qquad \ldots (2.11)$$

we arrive at the following lemma.

Lemma. If the variables $z_1, \ldots\ldots z_k$ and c_{ij} $(i,j=1, 2,\ldots k)$ are distributed as

$$\text{Const.}\ e^{-\frac{1}{2}\sum(z_i^2+c_{ii})+f\,z_1}\ |\,c_{ij}\,|^{\frac{q}{2}-\frac{k+1}{2}}\ \pi dz_i\ \pi\pi dc_{ij} \qquad \ldots (2.12)$$

then

(*i*) the statistic $V_r=\sum_1^r\sum_1^r c^{pq}_{(r)}\, z_p z_q$ is distributed as

$$\text{Const.}\ \frac{V_r^{\frac{r}{2}-1}}{(1+V_r)^{\frac{q+1}{2}}}\ {}_1F_1\left((q+1)/2\ ,\ r/2\ ,\ 2\,f^2\,\frac{V_r}{1+V_r}\right)dV_r \qquad \ldots (2.13)$$

so that when $f=0$ the statistic $V_r(q+1-r)/r$ can be used as the variance ratio with r and $(q+1-r)$ degrees of freedom and

(*ii*) the statistic $U=[(1+V_1)/(1+V_k)]-1$ is distributed as

$$\text{Const.}\ \frac{U^{\frac{k-r}{2}-1}}{(1+U)^{\frac{q+1-r}{2}}}\,dU \qquad \ldots (2.14)$$

so that $U(q+1-k)/(k-r)$ can be used as a variance ratio with $(k-r)$ and $(q+1-k)$ degrees of freedom.

3. GENERALISATION OF STUDENT'S t

Students' test connected with pairs of observations admits generalisation in two directions.

The first is to test whether the means of p correlated variates are the same on the basis of a sample of size n.

If $x_{1i}, \ldots\ldots x_{pi}$ are the observations on the variates corresponding to the i-th sample we replace the observations by a linear compound $x_i=l_1x_{1i}+\cdots\cdots+l_px_{pi}$ subject to the restriction $l_1+l_2+\ldots+l_p=0$. The problem, formally, reduces to testing whether the mean value of the variate x is zero. The appropriate statistic for this is $v=\sqrt{n}\bar{x}/s$ where

$$\left.\begin{aligned} \bar{x} &= l_1\,\bar{x}_1+\cdots\cdots+l_p\bar{x}_p \\ \bar{x}_j &= (x_{11}+\cdots\cdots+x_{jn})/n \\ (n-1)s^2 &= \textstyle\sum\sum l_i\,l_j\,b_{ij} \\ b_{ij} &= \textstyle\sum\,(x_{ir}-\bar{x}_i)(x_{jr}-\bar{x}_j) \end{aligned}\right\} \qquad \ldots (3.1)$$

The compounding coefficients $l_1, \ldots\ldots l_p$ are chosen so as to maximise this statistic or a constant times this statistic. Denoting the maximum value of $n\bar{x}^2/(n-1)s^2$ by V_{p-1} we get V_{p-1} as the root of the determinantal equation.

$$\begin{vmatrix} n\bar{x}_1\bar{x}_1 - V_{p-1}b_{11} & \ldots & n\bar{x}_1\bar{x}_p - V_{p-1}b_{1p} & 1 \\ \ldots & \ldots & \ldots & \vdots \\ n\bar{x}_p\bar{x}_1 - V_{p-1}b_{p1} & \ldots & n\bar{x}_p\bar{x}_p - V_{p-1}b_{pp} & 1 \\ 1 & \ldots & 1 & 0 \end{vmatrix} = 0 \qquad \ldots (3.2)$$

To find the value of V_{p-1} we may follow an alternative procedure which leads to the problem of distribution as well. By arbitrary choice of constants we can construct $(p-1)$ linear combinations

$$y_i = m_{i1}x_1 + \ldots + m_{ip}x_p \qquad \ldots (3.3)$$

such that $\Sigma m_{ij}=0$ for all $i=1, 2, \ldots\ldots (p-1)$. Choosing a linear compound $l_1x_1+\ldots+l_px_p$ of $x_1, x_2 \ldots x_p$ such that $\Sigma l=0$ is same as choosing an arbitrary linear compound of $y_1, \ldots y_{p-1}$ as defined in (3.3). If we choose the linear compound $\lambda_1 y_1+\ldots+\lambda_{p-1}y_{p-1}$ where λ's are free and construct the statistic

$$v = \frac{\lambda_1\bar{y}_1 + \ldots + \lambda_{p-1}\bar{y}_{p-1}\sqrt{n}}{\sqrt{\Sigma\,\Sigma\,\lambda_i\lambda_j c_{ij}}} \qquad \ldots (3.4)$$

where $c_{ij}=\Sigma(y_{ir}-\bar{y}_i)(y_{jr}-\bar{y}_j)$ we get the maximum value V'_{p-1} of v^2 as the root of the equation

$$| n\bar{y}_i\bar{y}_j - V'_{p-1}c_{ij} | = 0, \quad i, j, = 1, 2, \ldots (p-1) \qquad \ldots (3.5)$$

or

$$V'_{p-1} = \sum_1^{p-1}\sum_1^{p-1} n\, c^{ij}\, \bar{y}_i\, \bar{y}_j \qquad \ldots (3.6)$$

where c^{ij} are the elements of the matrix reciprocal to (c_{ij}), $i, j=1, 2 \ldots (p-1)$. The maximised values V_{p-1} and V'_{p-1} in the two cases must necessarily be the same so that $V_{p-1}=V'_{p-1}$. It immediately follows that the statistic V_{p-1} or V'_{p-1} is invariant for any system of parameters m_{ij} such that $\sum_j m_{ij}=0$ chosen to construct the $(p-1)$ variates $y_1, \ldots\ldots y_{p-1}$. In general we may choose a set which introduces simplicity in the evaluation of the statistic V_{p-1}.

The distribution problem is also simplified for we need only find the distribution of V_{p-1} from the joint distribution of $\bar{y}$'s and c_{ij}. As the statistic is invariant under linear transformations of the variates we can assume without loss of generality that the variates are distributed independently with unit variances and that the mean values of all y's except y_1 are zeroes. The mean value of y_1 denoted by m is zero on the null hypothesis. The actual value of m in terms of $m_1, \ldots m_p$ the mean values of $x_1, \ldots x_p$ and their variances and co-variances α_{ij} is given by the root of the equation.

$$\begin{vmatrix} m_i m_j - m^2\alpha_{ij} & \vdots & 1 \\ \ldots\ldots\ldots & \vdots & \ldots \\ 1 & \vdots & 0 \end{vmatrix} = 0 \qquad \ldots (3.7)$$

Since the probability density $\bar{y}$'s and c_{ij} is given by

$$\text{Const. } e^{-\frac{n}{2}\Sigma(\bar{y}_i^2 + c_{ii}) + n\, m\, \bar{y}_1} \mid c_{ij} \mid^{\frac{n-1}{2} - \frac{p}{2}} \qquad .. \quad (3.8)$$

We get the distribution of V_{p-1} as (2·13) with $r=p-1$, $q=n-1$ and $f=\sqrt{n}\, m$, so that to test the null hypothesis $m=0$ we may use the statistic $V_{p-1}(n-p+1)/(p-1)$ as the variance ratio with $(p-1)$ and $(n-p+1)$ degrees of freedom. It is also instructive to ascertain the values of $l_1, \ldots l_p$ of the compounding coefficients of the original variables so that we have a knowledge of a contrast leading to maximum discrepancy. The coefficients $l_1 \ldots l_p$ are obtainable from the linear equations

$$l_1 (n\, \bar{x}_1\, \bar{x}_i - V_{p-1}\, b_{1i}) + \ldots . + l_p (n\bar{x}_p\bar{x}_i - V_{p-1} b_{pi}) + \mu = o, \quad i = 1, 2, \ldots . p \qquad .. \quad (3.9)$$

and

$$l_1 + l_2 + \ldots . + l_p = o$$

where μ is also a variable to be solved for simultaneously.

On the other hand we can also test for the significance of t specified contrasts

$$y_i = l_{1i}\, x_1 + \ldots . + l_{pi}\, x_p, \quad \sum_j l_{ji} = o\,, \quad i = 1, 2 \ldots . t \qquad .. \quad (3.10)$$

We need only take a linear compound of y's and maximise a certain statistic. If we denote by

$$d^{ij} = \Sigma\, (y_{ir} - \bar{y}_i)\, (y_{jr} - \bar{y}_j) \qquad .. \quad (3.11)$$

then the appropriate statistic is

$$V_t = n\, \Sigma\, \Sigma\, d^{ij}\, \bar{y}_i\, \bar{y}_j \qquad .. \quad (3.12)$$

where d^{ij} are the elements of the matrix reciprocal to (d_{ij}), $i, j=1, \ldots t$. The quantities d_{ij} are derivable as linear combinations of b_{ij}'s defined in (3.1). To test for the significance of V_t we use the statistic $V_t(n-t)/t$ as the variance ratio with t and $(n-t)$ degrees of freedom.

If the further question is asked as to whether the contrasts (3.10) explain away the the differences among the p variates we have to compare V_t and V_{p-1} as derived from specified contrasts and all the available contrasts respectively. The distribution derived in (2.14) gives that the statistic

$$\frac{n-p+1}{p-t-1} \left(\frac{1+V_t}{1+V_{p-1}} - 1 \right) \qquad .. \quad (3.13)$$

can be used as a variance ratio with $(p-t-1)$ and $(n-p+1)$ degrees of freedom.

As an illustrative example we may consider the following problem connected with the testing of bias in using small sample plots in crop-cutting experiments.

The design, due to P. C. Mahalanobis, consisted in locating a random point in a field and constructing four concentric circles of radii 2 ft, 4 ft., 6 ft., and 8 ft. respectively. The inner circle is harvested first and the yield is recorded. The first annular, the second and third annular rings are harvested and the yields are separately recorded. From these by suitable addition we can get the yields as given by circular sample cuts of radii 2ft, 4 ft, 6 ft, and 8 ft at a chosen point. The yield rates (in some unit of weight per unit area) as given by the four circles are represented by c_1, c_2, c_3 and c_4 respectively. These variables are correlated and are subject to different errors depending on the correlation within a field. The problem is to test whether the mean yield rates as given by sample cuts of various sizes are the same.

Two dimensional charts representing the scatter of any pair of yield rates such as c_1 and c_2 gave that the arrays of c_2 for given c_1 are not homoscedastic and variance increases with increase in c_1. The data are then split into groups by the yield rate determined by c_1, so that within a group the arrays are nearly homoscedastic and the test for bias is carried on in each group. Incidentally the nature of bias for different intensities of yield rates can be studied.

In this case we may consider the variables

$$y_1=c_2-c_1,\ y_2=c_3-c_2,\ y_3=c_4-c_3$$

The mean values and the d_{ij} matrix constructed from 38 observations in the range of yield rate 0 to 10, are given below

$$\bar{y}_1 = -\cdot 1\ ,\ \bar{y}_2 = \cdot 43,\ \bar{y}_3 = \cdot 24$$

$$d_{ij} = \begin{pmatrix} 90\cdot 49 & -23\cdot 56 & -14\cdot 06 \\ -23\cdot 56 & 80\cdot 26 & -45\cdot 98 \\ -14\cdot 06 & -45\cdot 98 & 93\cdot 94 \end{pmatrix}$$

$$V_3 = \cdot 23$$

so that $35V_3/3=2\cdot 68$ can be considered as a variance ratio with 3 and 35 degrees of freedom. The result is not significant showing that in this group of yield rates there is no evidence of bias.

The second generalisation of students' t is concerned with testing, on the basis of a sample of size n from a $2p$ variate population containing the variables $y_1, y_2 \ldots y_{2p}$ whether $E(y_i)=E(y_{i+p})$ for $i=1, 2 \ldots p$.

From the $2p$ variates we construct the p variates $z_i=y_i-y_{i+p} (i=1, 2 \ldots p)$ in which case the test reduces to that of testing whether p correlated variables have assigned mean values viz, zeroes in this case ; a problem which has been considered by Hotelling (1935). We use the statistic

$$\frac{n-p}{p}\ V_p = \frac{n-p}{p}\ n\, \Sigma\, \Sigma\, \bar{z}_i\, \bar{z}_p\, d^{ij} \qquad \ldots (3.14)$$

as the variance ratio with p and $(n-p)$ degrees of freedom.

The problem where the p variates of the first set are uncorrelated with the variates of the second set and the dispersion matrices of the two sets are identical can be answered with the use of Mahalanobis' Generalised distance.

In a special case where the estimates of variances and covariances come out as constant multipliers of a stochastic variable the problem can be answered with the use of Mahalanobis' Generalised distance but the distribution in this case is different from the one to be used above. This, statistic has been termed by the author as the Mahalanobis' distance of the second kind in (Rao : 1944). The class of hypotheses arising out of these problems can be appropriately treated as tests in least squares fully discussed by the author in (Rao : 1946).

4. A PROBLEM IN THE CLASSIFICATION OF THREE POPULATIONS

An important problem that arises in the classification of three multivariate populations π_1 π_2 and π_3 is to test whether the population π_2 is nearer to one of π_1 and π_3 when it is known that π_1 and π_3 are different. If the variates corresponding to the three populations π_1, π_2 and π_3 are represented by $x_1, \ldots x_p$; $y_1, \ldots y_p$; and $z_1, \ldots z_p$ then we replace the p variates by a linear combination of these variates defined by $\mathrm{x}=\Sigma lx$, $\mathrm{y}=\Sigma ly$, $\mathrm{z}=\Sigma lz$. The problem is thus formally reduced to the case of a univariate classification.

There are two contrasts arising out of the three mean values of the variables x,y and z which give the inequalities in the mean values of the three populations. One such contrast is supplied by $\mathrm{E(x)}=\mathrm{E(z)}$ which gives the differences in the mean values of π_1 and π_3. Another contrast is $2\mathrm{E(y)}-\mathrm{E(x)}-\mathrm{E(z)}$ which determines the nearness of π_2 to one of π_1 and π_3. We take this contrast and choose the compounding coefficients so as to maximise the ratio of its estimate to the standard error.

If n_1, n_2 and n_3 are sample sizes available for the populations π_1, π_2 and π_3 then assuming equality of variances and covariances we can build up the estimates of variances and covariances from the quantities

$$c_{ij} = \sum_{r=1}^{n_1} (x_{ir}-\bar{x}_i)\, x_{jr} + \sum_{r=1}^{n_2} (y_{ir}-\bar{y}_i)\, y_{jr} + \sum_{r=1}^{n_3} (z_{ir}-\bar{z}_i)\, z_{jr} \qquad \ldots (4.1)$$

The ratio to be maximised is

$$\frac{\sqrt{n}\ \ \Sigma\, l_i\,(2\bar{y}_i-\bar{x}_i-\bar{z}_i)}{\sqrt{\Sigma\,\Sigma\, l_i\, l_j\, c_{ij}}} \qquad \ldots (4.2)$$

where $1/\mathrm{n}=4/n_2+1/n_1+1/n_3$. The maximum value of the square of (4.2) comes out as

$$\mathrm{V}_p = \sqrt{\mathrm{n}}\ \Sigma\,\Sigma\, c^{ij}\,(2\,\bar{y}_i-\bar{x}_i-\bar{z}_i)(2\bar{y}_j-\bar{x}_j-\bar{z}_j) \qquad \ldots (4.3)$$

where c^{ij} are the elements of the matrix reciprocal to (c_{ij}). Since c_{ij} are determined with $(n_1+n_2+n_3-3)$ degrees of freedom it follows that the statistic

$$\frac{n_1+n_2+n_3-2-p}{p}\ \mathrm{V}_p \qquad \ldots (4.4)$$

can be used as the variance ratio with p and $(n_1+n_2+n_3-p-2)$ degrees of freedom.

When the test gives a significant result the nearness to π_1, or π_3 has to be determined by the evaluation of Mahalanobis' generalised distances between π_1, π_2 and π_3, π_2. If the former is greater than the latter then π_2 is nearer to π_3 and viceversa.

This test can be extended to answer another type of problem considered by Wald (1944) connected with the classification of an individual into one of two groups, π_1 and π_3 from which samples of sizes n_1 and n_3 are available. We need construct the statistic (4.3) with $n_2=1$ and test for the significance of V_p. If V_p is significant then the individual can be classified as belonging to π_1 or π_3. If not we may have to be doubtful about the classification of the individual. This appears to be a symmetrical test when the individual is known to belong to one of the two groups. The exact nature of this test, however, requires further investigation. The same statistic (4·3) cad be used when a group of n_2 individuals are to be classified with one of two groups π_1 and π_3 from which samples of sizes n_1 and n_3 are available.

REFERENCES

1. BARTLET, M. S. (1939). The standard errors of discriminant function coefficients. *J. R. S. S. Suppl.* 6, 169-73.
2. FISHER, R. A. (1938). The statistical utilisation of multiple measurements. *Ann. of Engen.*, 8, 376-86.
3. ——— (1940). The precision of discriminant functions. *Ann. of Engen.*, 10, 422-29.
4. RAO, C. R. (1944) Generalised variance of populations *Proc Ind. Sc. Cong.*
5. ——— (1946). On linear combinations of observations and the general theory of least squares, *Sankhya*, 7.
6. WALD, A. (1945). On a statistical problem arising in the classification of an individual in to one of two groups. *Ann. of Math. Stat.* 15, 145-62.

Paper received : 9 September 1945

CONFIDENCE INTERVAL FOR THE CORRELATION COEFFICIENT

By K. C. S. PILLAI

Travancore University, Travancore.

INTRODUCTION

The distribution of the correlation coefficient r, when the population value ρ is zero was for the first time studied by Student (1903). His method was however purely empirical. The exact distribution of r was subsequently given by Fisher (1915) based on geometrical methods. A co-operative study (1917) was undertaken in the Galton Laboratory regarding the sampling distribution of r. The result of these studies gave extensive tables of the ordinates of the curve. Recently Miss David has tabled the probability integral of r for different values of ρ and n the sample size. These tables have to a large extent simplified the practical application of Fisher's exact distribution.

Fisher (1921) has effected a logarithmic transformation of r which in large samples reduces the distribution of r to the normal probability scale. This is indeed a great step and tests of significance of r have become fairly simple. Miss David's tables together with Fisher's transformation are in general sufficient to tackle most of the problems relating to the sampling distribution of r.

The problem of finding unbiased confidence interval for r in the sense of Neyman-Pearson (1936) has not been given an elegant solution, though in any particular case it is possible to set up these intervals by trial. After examining a few cases Miss David (1937) has come to the conclusion that the unbiased limits for r differ very little from those corresponding to equal tail areas. A detailed examination of different methods of arriving at confidence limits is attempted in this paper.

2. THE PROBLEM OF CONFIDENCE INTERVALS

The problem of confidence intervals may be briefly outlined here. Consider the probability law $P(x, \theta,)$, θ being an unknown parameter. Let $x_1, x_2, \ldots x_n$ be a random sample and let T be a sufficient statistic for θ, so that the sampling distribution of T involes θ. To test the hypothesis $\theta=\theta_0$ based on evidence available in the sample, it is clear that two numbers $A(\theta_0)$ and $B(\theta_0)$ have to be determined such that

$$\int_{A(\theta_0)}^{B(\theta_0)} P(T,\theta_0)\, dT=\alpha(\text{say } 0{\cdot}95) \qquad .. \quad (2.1)$$

Since equation (2.1) cannot uniquely determine $A(\theta_0)$ and $B(\theta_0)$, a second condition has to be laid down. Amongst different methods of specifying this condition, the following three are important.

(1) *Equal tail areas.* Here it is assumed that .. (2.2)

$$P[T \leqslant A(\theta_0)] = P[B(\theta_0) \leq T] = (1-\alpha)/2$$

(2) *Shortest Confidence Interval.* Here $A(\theta_0)$ and $B(\theta_0)$ are determined from the equations. (2.3)

$$P[A(\theta) \leqslant T \leqslant B(\theta)]=\alpha \text{ and } B(\theta)-A(\theta) \text{ is a minimum for } \theta=\theta_0 \; \cdot\cdot$$

(3) *Neyman-Pearson's method of unbiased region of acceptance.* Denote by $\beta(\theta)$ the probability that T lies outside the interval $[A(\theta), B(\theta)]$ so that

$$\beta(\theta) = 1 - \int_{A(\theta)}^{B(\theta)} P(T, \theta)\, dT \qquad \cdot\cdot \quad (2.4)$$

In testing the hypothesis $\theta=\theta_0$, the rule of acceptance if T lies in the interval $[A(\theta_0), B(\theta_0)]$ gives $\beta(\theta_0)$ as the probability of rejection of the hypothesis. This has been called the first kind of error and by taking α near unity the first kind of error may be kept at any desired low level. Clearly the test has to be so designed as to throw out the hypothesis $\theta=\theta_0$ whenever some other alternative value of $\theta\neq\theta_0$ is true. The probability of rejection of $\theta=\theta_0$ when it is not true is given by

$$\gamma(\theta) = 1 - \int_{A(\theta_0)}^{B(\theta_0)} P\,(T,\theta)\, dT \qquad \cdot\cdot \quad (2.5)$$

This probability has therefore to be a minimum at $\theta=\theta_0$. Thus $A(\theta_0)$ and $B(\theta_0)$ are obtained by the relations

$$\beta(\theta_0) = \gamma(\theta_0) = 1-\alpha \text{ and } \left.\frac{d\gamma}{d\theta}\right|_{\theta=\theta_0} = 0 \qquad \cdot\cdot \quad (2.6)$$

The interval $[A(\theta_0), B(\theta_0)]$ satisfying (2.6) is called unbiased confidence region of acceptance and $\gamma(\theta)$ is called the power of the test.

The confidence limits for the correlation coefficient r are discussed in the light of the three methods mentioned above.

3. CONFIDENCE LIMITS BASED ON FISHER'S Z-TRANSFORMATION

The exact distribution of the correlation coefficient r given by Fisher is of the form

$$P(r\,|\,n,\rho) = \frac{(1-\rho^2)^{(n-1)/2}}{\pi\Gamma(n-2)} (1-r^2)^{(n-4)/2} \cdot \frac{d^{n-2}}{d(r\rho)^{n-2}} \frac{\text{arc cos }(-\rho r)}{\sqrt{1-\rho^2r^2}} \qquad \cdot\cdot \quad (3.1)$$

Fisher's z-transformation is

$$z = (\sqrt{n-3}/2) \log [(1+r)(1-\rho)/(1-r)(1+\rho)] \qquad \cdot\cdot \quad (3.2)$$

and z is distributed normally for large n and ρ small, round zero with unit standard deviation. Hence if a is detremined such that

$$\frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} exp(-z^2/2)dz = \alpha \quad (\text{say } 0.05) \qquad \cdot\cdot \quad (3.3)$$

the probability for the inequality

$$-a \leq (\sqrt{n-3}/2) \log [(1+r)(1-\rho)/(1-r)(1+\rho)] \leq a \qquad \cdot\cdot \quad (3.4)$$

The inquality (3.4) is equivalent to $A(\rho) \leq r \leq B(\rho)$ where

$$\left.\begin{aligned} A(\rho) &= [(1+\rho)\exp(-2a/\sqrt{n-3}) - (1-\rho)]/[(1+\rho)\exp(-2a/\sqrt{n-3}) + (1-\rho)] \\ B(\rho) &= [(1+\rho)\exp(2a/\sqrt{n-3}) - (1-\rho)]/[(1+\rho)\exp(2a/\sqrt{n-3}) + (1-\rho)] \end{aligned}\right\} \qquad \cdot\cdot \quad (3.5)$$

To test the hypothesis $\rho=\rho_0$ the acceptance region is given by $[A(\rho_0), B(\rho_0)]$. It is obvious that $P[r \leqslant A(\rho_0)] = P(z \leqslant -a)$ and $P[B(\rho_0) \leqslant r] = P(a \leqslant z)$. Hence in virtue of (3.3) the region of acceptance $[A(\rho_0), B(\rho_0)]$ corresponds to that of the method of equal tail areas. To examine whether this region satisfies the condition of unbiasedness in the sense of Neyman-Pearson, the value of $\gamma(\rho)$ in (2.5) has to be evaluated. It will be observed that in view of the transformation (3.2) the inequality $[A(\rho_0) \leqslant r \leqslant B(\rho_0)]$ reduces to

$$\delta = -a + (\sqrt{n-3}/2) \log [(1+\rho_0)(1-\rho)/(1-\rho_0)(1+\rho)] \leqslant z \qquad \cdot\cdot \quad (3.6)$$

$$\leqslant + a\,(\sqrt{n-3}/2)\log[(1+\rho_0)(1-\rho)/(1-\rho_0)(1+\rho)] = \mu \quad \text{.. (3.7)}$$

Hence

$$\gamma(\rho) = 1-(1/\sqrt{2\pi})\int_0^{\mu} e^{-z^2/2}\,dz \quad \text{.. (3.8)}$$

and

$$\frac{d\gamma}{d\rho} = (1/\sqrt{2\pi})[e^{-\mu^2/2}\frac{d\mu}{d\rho} - e^{-\delta^2/2}\frac{d\delta}{d\rho}] \quad \text{... (3.9)}$$

which gives that $d\gamma/d\rho=0$ when $\rho=\rho_0$. Thus the test is unbiased. However, since the transformation of r to z makes the distribution of z tend to normality when n is large the above theory cannot be expected to make the test unbiased strictly. In fact, from the tabled distribution of r by David $\gamma(\rho)=1-\mathrm{P}[\mathrm{A}(\rho_0)\leqslant r\leqslant \mathrm{B}(\rho_0)]$ can be evaluated and this shows that there is a bias in the test though it is negligible. The power function of the test is tabled in Section 6.

4. Nair's transformation of the correlation coefficient

It is interesting to study the transformation,

$$x = (\rho - r)\,/\,(1-\rho r) \quad \text{.. (4.1)}$$

suggested by Dr. U. S. Nair, which appears to yield the confidence limits for r in a simple manner. In the first place, the relation between this transformation and Fisher's z is

$$z = (\sqrt{n-3}/2)\log[(1+x)/(1-x)] \quad \text{.. (4.2)}$$

which is independent of ρ. Secondly it will be noted that

$$\begin{aligned}\mathrm{P}[(\rho-a)/(1-\rho a) &\leqslant r \leqslant (\rho+a)/(1+\rho a)]\\ &= \mathrm{P}(-a\leqslant x\leqslant a)\end{aligned} \quad \text{.. (4.3)}$$

for any fixed value of $a<1$. The sampling distribution of x is studied below. It is seen that $\mathrm{P}(x)$ is independent of ρ approximately. Hence the probability given in (4.3) is independent of ρ and consequently,

$$\mathrm{A}(\rho_0) = (\rho_0-a)\,/\,(1-\rho_0 a) \text{ and } \mathrm{B}(\rho_0) = (\rho_0+a)\,/\,(1+\rho_0 a) \quad \text{.. (4.4)}$$

Getting a such that the probability in (4.3) is α, unbiased limits for r are obtained.

It now remains to find the sampling distribution of x. In a Co-operative Study (1917) it is shown that

$$\begin{aligned}\mathrm{P}(r) &= [n-2)/(\sqrt{2\pi}\ \overline{n-1})][(1-\rho^2)^{(n-1)/2}(1-r)^{(n-4)/2}/(1-\rho r)^{(2n-3)/2}]\\ &\times\{1+[(\rho r+2)/8(n-1)]+[(3\rho r+2)^2/128(n-1)^2]+\ldots\}\end{aligned} \quad \text{.. (4.5)}$$

Substituting for r from (4.1) and using $\mathrm{P}(r)dr=\mathrm{P}(x)dx$

$$\begin{aligned}\mathrm{P}(x) &= [(n-2)/\sqrt{2\pi}\ \overline{(n-1)\ (1-\rho^2)}][(1-x^2)^{(n-4)/2}(1-\rho x)^{1/2}]\\ &\times\ \{1+(2+\rho^2-3\rho x)/8(n-1)(1-\rho x)+(2+3\rho^2-5\rho x)^2/128(n-1)^2(1-\rho x)^2+\ldots\ldots\}\end{aligned} \quad \text{.. (4.6)}$$

Neglecting the terms $0(1/n)$ .. (4.7)

$$\mathrm{P}(x)\sim[(n-2)/\sqrt{2\pi(n-1)(1-\rho^2)}](1-x^2)^{(n-4)/2}(1-\rho x)^{1/2}$$

which may be reduced to

$$\begin{aligned}\mathrm{P}(x)\sim\{\Gamma[(n-1)/2]/\sqrt{\pi}\,\Gamma[(n-2)/2]\}(1-x^2)^{(n-4)/2} &\times [1-\rho x/2+ \text{ terms involving} && \text{.. (4.8)}\\ &\quad \rho^2 \text{ and } 1/n] && \text{.. (4.9)}\end{aligned}$$

Thus $\mathrm{P}(x)\sim\{\Gamma[(n-1)/2]/\sqrt{\pi}\,\Gamma[(n-2)/2]\}(1-x^2)^{(n-4)/2}$ However confining to the problem
approximately, omitting terms $0(1/n)$ and powers of $\rho\geqslant 1$. since the term $-\rho x/2$ in (4.8)
of evaluating the probability integral for r from $\mathrm{P}(-a\leqslant x\leqslant a)$ is valid if n be large and ρ^2 negli-
does not alter the value of the integral, approximation (4.9)
gible.

To compare the accuracy of (4.9) in determining the probability levels for r, in small samples, the following calculations have been made.

The transformation (4.1) gives $P(-1 \leq r \leq \mu \mid \rho) = P(\gamma \leq x \leq 1)$ where $\gamma = (\rho - \mu)/(1 - \rho\mu)$

$$\left.\begin{aligned} &= 1/2 - 1/2\, I_{\gamma^2}(1/2, \overline{n-2}/2) \text{ if } \gamma \text{ is positive} \\ &= 1/2 + 1/2\, I_{\gamma^2}(1/2. \overline{n-2}/2) \text{ if } \gamma \text{ is negative} \end{aligned}\right\} \quad \ldots \quad (4.10)$$

$$\text{Where } I_{\gamma_2}(1/2, \overline{n-2}/2) = \{\Gamma[(n-1)/2]/\Gamma[(n-2)/2]\sqrt{\pi}\} \int_0^{\gamma^2} t^{-1/2}(1-t)^{(n-4)/2} dt \quad .. \quad (4.11)$$

From Tables of Incomplete Beta Function, the value of (4.10) can be calculated for different values of n, ρ and r. Also Miss David's tables will give the correct value of $P(-1 \leq r \leq \mu \mid \rho)$. The following table gives the comparison of these two calculations.

TABLE 1. VALUES OF $P\ (-1 \leq r \leq \eta \mid \rho)$

μ	$n = 5,\ \rho = 0{\cdot}1$		$n = 50,\ \rho = 0{\cdot}8$	
	David's Table	Nair's transform	David's Table	Nair's transform
—1 ·0	0 ·0000	0 ·0000	0 ·0000	0 ·0000
—0 ·8	0 ·0379	0 ·0398	0 ·0010	0 ·0023
—0 ·6	0 ·1077	0 ·1125	0 ·0035	0 ·0075
—0 ·2	0 ·3056	0 ·3155	0 ·0153	0 ·0301
0 ·0	0 ·4253	0 ·4364	0 ·0280	0 ·0520
0 ·3	0 ·6197	0 ·6303	0 ·0681	0 ·1137
0 ·6	0 ·8151	0 ·8219	0 ·1826	0 ·2613
0 ·9	0 ·9740	0 ·9723	0 ·6505	0 ·7224
1 ·0	1 ·0000	1 ·0000	1 ·0000	1 ·0000
	$n=10,\ \rho=0{\cdot}1$		$n=10,\ \rho=0{\cdot}8$	
—1 ·0	0 ·0000	0 ·0000	0 ·0000	0 ·0000
—0 ·8	0 ·0013	0 ·0014	0 ·0000	0 ·0000
—0 ·6	0 ·0181	0 ·0177	0 ·0000	0 ·0000
—0 ·2	0 ·1998	0 ·2047	0 ·0004	0 ·0007
0 ·0	0 ·3849	0 ·3917	0 ·0016	0 ·0027
0 ·3	0 ·7103	0 ·7162	0 ·0125	0 ·0188
0 ·6	0 ·9414	0 ·9484	0 ·1037	0 ·1362
0 ·9	0 ·9996	0 ·9995	0 ·8135	0 ·8445
1 ·0	1 ·0000	1 ·0000	1 ·0000	1 ·0000
	$n = 15,\ \rho = 0{\cdot}1$		$n = 15,\ \rho = 0{\cdot}8$	
—1 ·0	0 ·0000	0 ·0000	0 ·0000	0 ·0000
—0 ·8	0 ·0001	0 ·0001	0 ·0000	0 ·0000
—0 ·6	0 ·0036	0 ·0037	0 ·0000	0 ·0000
—0 ·2	0 ·1406	0 ·1437	0 ·0000	0 ·0000
0 ·0	0 ·3563	0 ·3615	0 ·0001	0 ·0002
0 ·3	0 ·7654	0 ·7696	0 ·0025	0 ·0038
0 ·6	0 ·9787	0 ·9794	0 ·0614	0 ·0785
0 ·9	1 ·0000	1 ·0000	0 ·8872	0 ·9044
1 ·0	1 ·0000	1 ·0000	1 ·0000	1 ·0000

The calculations have not been made for $n > 15$ in the above table as this has been done in greater detail in Table 2. It is evident from the above table that the approximation for small samples is particularly bad for $n=5$ and $\rho=0.8$. However for $n>10$ the approximation considerably improves even for $\rho=0.8$. For small values of ρ(say 0·1) for $n=10$ and higher values the approximation is more or less satisfactory.

In spite of the inaccuracies noticeable in Table 1 it is interesting to note that $P[(\rho-a)/(1-\rho a) \leq r \leq (\rho+a)/(1+\rho a)]$ taken from Miss David's Tables is in fair agreement with its value given by the approximation. Of course for $n=5$ and 10 the approximation is good only for small values of ρ. But for n larger than 10 (say 15, 25, etc.) even for high values of ρ the approximation is good. The accuracy of the approximation is also compared with that based on Fisher's z-transformation.

Table 2 gives values of I_1, I_2, I_3, where $I_1 = P[A(\rho_0) \leq r \leq B(\rho_0)]$ from David's Tables.
$I_2 =$ Value of I_1 from the approximation based on the transformation $r = (\rho - x)/(1 - \rho x)$
$I_3 =$ Value of I_1 from the approximation based on the transformation

$$z = (\sqrt{n-3}/2) \log [(1+r)(1-\rho)/(1-r)(1+\rho)]$$

which gives
$I_3 = P[(\sqrt{n-3}/2) \log (1-a)/(1+a) \leq z \leq (\sqrt{n-3}/2) \log (1+a)/(1-a)]$, z being a normal variate, $A(\rho_0)$ and $B(\rho_0)$ being given by (4.9).

TABLE 2. ON THE ACCURACY OF THE CONFIDENCE LIMITS BASED ON A (ρ_0) and B (ρ_0)

				$P[A(\rho_0) \leqslant r \leqslant B(\rho_0)]$		
				I_1	I_2	I_3
	ρ_0	$A(\rho_0)$	$B(\rho_0)$	(true value)		
n=5 a=0.2	0·4	0·21739	0·55556	0·2543	0·2559	0·2257
	0·8	0·71429	0·86207	0·2588	0·2529	0·2257
n=5 a=0·6	0·2	—0·45455	0·71129	0·7161	0·7152	0·6730
	0·7	0·17241	0·91549	0·7247	0·7152	0·6730
	0·9	0·65217	0·97403	0·7287	0·7152	0·6730
n=10 a=0·1	0·3	0·20619	0·38835	0·2169	0·2166	0·2093
	0·9	0·87912	0·91743	0·2192	0·2166	0·2093
n=10 a=0·7	0·2	—0·58140	0·78947	0·9758	0·9758	0·9780
	0·6	—0·17241	0·91549	0·9763	0·9758	0·9780
	0·8	0·22727	0·96154	0·9762	0·9758	0·9780
n=15 a=0·4	0·4	0·00000	0·68965	0·8609	0·8504	0·8578
	0·9	0·78125	0·95588	0·8623	0·8604	0·8578
n=15 a=0·8	0·5	—0·50000	0·92857	0·9997	0·9997	0·9999
	0·8	0·00000	0·97561	0·9997	0·9997	0·9999
n=25 a=0·3	0·7	0·50633	0·82645	0·8559	0·8549	0·8534
n=25 a=0·9	0·9	0·00000	0·99448	1·0000	1·0000	1·0000

From this table it becomes evident that in small samples the transformation suggested in the paper is slightly better than that of Fisher. For $n<15$ the approximation appears to be adequate only for small values of ρ. For large values of n even when ρ is near unity, the approximation gives sufficiently accurate levels.

The accuracy of the transformation (4.1) justifies the use of $A(\rho_0)$ and $B(\rho_0)$ given in (4.4) as unbiased limits for r. The power function of this test is examined in detail in Section 6 along with that based on Fisher's z-transformation.

5. SHORTEST CONFIDENCE INTERVAL

Since the distribution of r is very complicated in form the theoretical method of determining $A(\rho_0)$ and $B(\rho_0)$ satisfying equation (2.3) is not attempted. A graphical method to determine the shortest confidence interval for a variate from a given frequency table has been suggested by Miss A. George (1942). Let n_i be the frequency corresponding to the value of x_i of x and Σn_i the total frequency. If $z_i=(n_1+\ldots+n_i)/N$ the curve (x_i, z_i) gives the cumulative frequency curve and the ordinate of this curve gives the probability integral of x.

Suppose a is a value of x. To find b such that $P(a\leq x\leq b)=\alpha$ draw the ordinate corresponding to a and produce it such that the total length of the line is the sum of the ordinate and α. If the line through its extremity parallel to the x-axis cuts the curve in some point, then the abscissa of that point gives the value of b.

The shortest confidence interval is obtained by minimising $(b-a)$. Take a set of values of a and find the corresponding values of $(b-a)$ from the cumulative curve. By plotting $(b-a)$ against a the value of a that minimises $(b-a)$ can be obtained graphically. This value gives the lower limit and the upper limit is obtained by increasing it by the $(b-a)$.

The shortest confidence intervals for r for the hypotheses tested according to the methods in the last Sections have been obtained with the help of the Tables of Correlation Coefficient. The power function of this test has also been calculated in Section 6.

6. POWER CURVES AND DISCUSSION

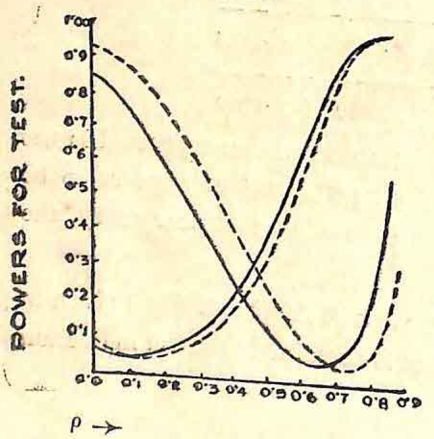
This section contains the calculation of the power function defined by (2.5) corresponding to the different methods discussed above. These are— 1. The new transformation of r suggested by Nair. 2. Fisher's z-transformation. 3. The method of Shortest Confidence Interval.

The tables give $1-P[(A(\rho_0)\leq r\leq B(\rho_0)]$. The power curves based on the values in Table 3 are also given.

POWER CURVES.

$(\rho_0=0{\cdot}1;\ 0{\cdot}7,\ \alpha=0{\cdot}05.)$

$n=15$



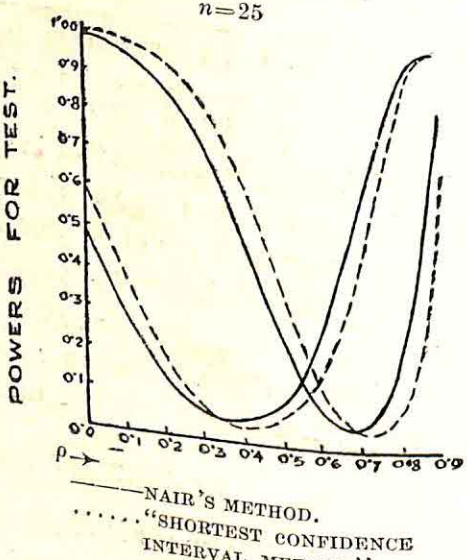
———NAIR'S METHOD.
......"SHORTEST CONFIDENCE INTERVAL METHOD".

FIG. 1.

POWER CURVES.

$(\rho_0=0{\cdot}4;\ 0{\cdot}7,\ \alpha=0{\cdot}05)$

$n=25$



———NAIR'S METHOD.
....."SHORTEST CONFIDENCE INTERVAL METHOD".

FIG. 2.

POWER CURVES.
$(\rho_0 = 0{\cdot}2;\ 0{\cdot}8,\ \alpha = 0{\cdot}05)$
$n = 50$

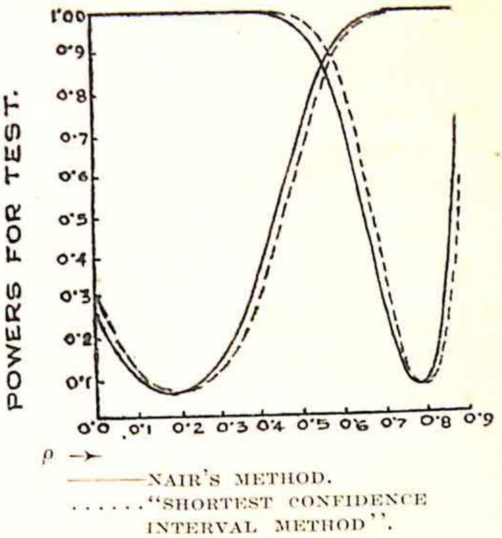


FIG. 3-

TABLE 3. PROBABILITY OF REJECTING H_0 $[\rho_0, n, \alpha = 0{\cdot}95]$.

			$n=15$			
ρ		$\rho_0 = 0{\cdot}1$		$\rho_0 = 0{\cdot}7$		
	Nair's method	Fisher's method	S.C.I. method	Nair's method	Fisher's method	S.C.I. method
0·0	0·063	0·064	0·069	0·85	0·86	0·92
0·1	0·050	0·051	0·050	0·75	0·76	0·86
0·2	0·066	0·067	0·059	0·62	0·63	0·76
0·3	0·115	0·117	0·101	0·47	0·47	0·62
0·4	0·210	0·213	0·187	0·31	0·31	0·46
0·5	0·362	0·365	0·330	0·17	0·17	0·29
0·6	0·568	0·571	0·533	0·08	0·08	0·14
0·7	0·790	0·792	0·763	0·05	0·05	0·05
0·8	0·948	0·949	0·939	0·14	0·14	0·05
0·9	0·998	0·998	0·997	0·60	0·60	0·34
$A(\rho_0)$	—0·43643	—0·43454	—0·42000	0·29053	0·29269	0·38900
$B(\rho_0)$	0·58398	0·58243	0·60000	0·89278	0·89230	0·92500
			$n=25$			
		$\rho_0 = 0{\cdot}4$		$\rho_0 = 0{\cdot}7$		
0·0	0·508	0·511	0·594	0·982	0·982	0·991
0·1	0·321	0·324	0·402	0·948	0·948	0·973
0·2	0·170	0·172	0·229	0·872	0·874	0·925
0·3	0·078	0·078	0·108	0·734	0·736	0·823
0·4	0·050	0·050	0·050	0·529	0·532	0·650
0·5	0·096	0·097	0·060	0·297	0·300	0·412
0·6	0·258	0·261	0·171	0·114	0·115	0·181
0·7	0·575	0·578	0·453	0·050	0·050	0·050
0·8	0·903	0·904	0·842	0·208	0·212	0·102
0·9	0·999	0·999	0·997	0·836	0·838	0·685
$A(\rho_0)$	0·00416	0·00577	0·05000	0·42010	0·42143	0·47250
$B(\rho_0)$	0·68747	0·68661	0·72000	0·85828	0·85786	0·88250
			$n=50$			
		$\rho_0 = 0{\cdot}2$		$\rho_0 = 0{\cdot}8$		
0·0	0·282	0·284	0·303	1·000	1·000	1·000
0·1	0·106	0·107	0·116	1·000	1·000	1·000
0·2	0·050	0·050	0·050	1·000	1·000	1·000
0·3	0·116	0·117	0·107	0·999	0·999	1·000
0·4	0·336	0·338	0·317	0·996	0·996	0·998
0·5	0·673	0·675	0·653	0·961	0·962	0·978
0·6	0·926	0·927	0·918	0·780	0·781	0·846
0·7	0·996	0·996	0·995	0·336	0·338	0·429
0·8	1·000	1·000	1·000	0·050	0·050	0·050
0·9	1·000	1·000	1·000	0·746	0·749	0·620
$A(\rho_0)$	—0·08367	—0·08297	—0·07500	0·67070	0·67109	0·69000
$B(\rho_0)$	0·45368	0·45313	0·46000	0·88211	0·88188	0·89300

It is obvious that there is not great perceptible difference between the power curves corresponding to Nair's and Fisher's transformations which one has to expect as a result of the relation (4.2) between x and z. The shortest confidence interval for r however shows bias. The bias becomes negligible as the sample size increases.

In conclusion the object of the paper has been to study different methods of setting up the confidence limits for r and examine the nature of the power curves in each case. A new transformation for r is suggested which in large samples reduces the distribution of r to symmetrical Beta-Function independent of ρ. This transformation and Fisher's transformation yield almost identical values for probability level of r in large samples. In small samples however, the approximation suggested in the paper appear to give closer values of probability than that of Fisher.

References

1. Student (1906) *Biometrika VI* 302-10.
2. R. A. Fisher. (1915) *Biometrika X* 507 21.
3. H. E. Soprr. (1913) *Biometrika IX* 91-115.
4. A Co-operative Study. (1917) *Biometrika XI* 328-413.
5. Miss David. Tables of Correlation Coefficient.
6. R- A. Fisher. (1921) *Metron 1, IV* 1-30.
7. J. Neyman and E. S. Prarson. (1936) *Statistical Research Memoirs, 1,* 1-37.
8. F. N. David. (1937) *Biometrika XXIX* 157-60.
9. Miss. A. George. (1942) On the problem of Interval Estimation. *Sankhya* 6, 2 Part 111-120.

Paper received : 31 January 1946.

A NEW SOLUTION IN SYMMETRICAL BALANCED INCOMPLETE BLOCK DESIGNS ($v=b=31$, $r=k=10$, $\lambda=3$)

By K. N, BHATTACHARYA

Calcutta University, Calcutta.

1. If v objects are arranged in b sets of k objects each, such that every object occurs in just r sets, and every pair of objects occurs in just λ sets, then we get a 'Balanced Incomplete Block Design'. Such designs are important for agricultural field experiments and other biological researches where statistical methods are applicable.

Following the terminology of agricultural field experiments, the v objects are called 'treatments' and the b sets are called blocks. Thus there are v treatments and b blocks. The parameters v, b, r, k, λ are obviously subject to the relations $bk=vr$, $\lambda(v-1)=r(k-1)$ and as shown by Fisher $b \geqslant v$ or $k \leqslant r$.

Only designs for which $r \leqslant 10$, are of practical interest. Integral values of the parameters consistent with the above were given in the first edition of Fisher and Yates' tables (1938), together with the known combinatorial solutions. A number of solutions for previously unknown cases were added later by R. C. Bose (1939) and others. These researches have enabled Fisher and Yates (1943) to fill in, in the second edition, some of the blanks which were left in the previous one. However the following 12 blanks were still left.

reference no. in Fisher and Yates' tables	parameters v	b	r	k	λ	reference no. in Fisher and Yates' tables	parameters v	b	r	k	λ
(8)	15	21	7	5	2	(24)	46	69	9	6	1
(10)	22	22	7	7	2	(26)	21	30	10	7	3
(12)	21	28	8	6	2	(27)	31	31	10	10	3
(14)	29	29	8	8	2	(28)	36	45	10	8	2
(17)	16	24	9	6	3	(30)	46	46	10	10	2
(20)	25	25	9	9	3	(31)	51	85	10	6	1

Mr. Q. M. Husain (1945, 1946) has demonstrated the impossibility of obtaining a solution for the designs no. (10) and (14), and Mr. H. K. Nandi (1945) demonstrated the same for the design no. (8). The author (1944, 1945) gave solutions for thedesigns number (17) and (20). The object ofthepresent note is to give a solution for the designs no. (26) and (27). This leaves only 5 cases to be accounted for, including no. (12), which Mr. Nandi conjectures to be impossible, though an exhaustive verification of this conjecture has not yet been carried out.

2. The solution for the symmetrical design no. (27) with parameters $v = b = 31$, $r = k = 7$, $\lambda = 3$ is as follows :

Consider the module of residue classes (mod 7). To each element u of the module let us associate four treatments u_1, u_2, u_3, and u_4. To the 28 treatments thus obtained, let us adjoin the three new treatments $\infty_1, \infty_2, \infty_3$. To obtain the solution for thedesign no. (27) we start with the four initial blocks,

B_o : $(1_1, 2_1, 4_1 ; 1_2, 2_2, 4_2 ; 1_3, 2_3, 4_3 ; 0_4)$, B_1 : $(1_1, 6_1 ; 2_2, 5_2 ; 3_3, 4_3 ; 3_4, 5_4, 6_4, \infty_1)$

B_2 : $(2_1, 5_1 ; 3_2, 4_2 ; 1_3, 6_3 ; 3_4, 5_4, 6_4, \infty_2)$, B_3 : $(3_1, 4_1 ; 1_2, 6_2 ; 2_3, 5_3 ; 3_4, 5_4, 6_4, \infty_3)$

and develop them (mod. 7). Each initial block thus gives rise to 7 blocks obtained from it, by successively adding the elements of the module, keeping ∞'s and the suffixes invariant. The solution is now completed by adding the three new blocks

B'_1 : $(0_1, 1_1, 2_1, 3_1, 4_1, 5_1, 6_1, \infty_1, \infty_2, \infty_3)$, B'_2 : $(0_2, 1_2, 2_2, 3_2, 4_2, 5_2, 6_2, \infty_1, \infty_2, \infty_3)$

B'_3 : $(0_3, 1_3, 2_3, 3_3, 4_3, 5_3, 6_3, \infty_1, \infty_2, \infty_3)$

3. The proof can easily be given following the ideas developed by Bose (1939, 1942) We can consider the treatments other than ∞_1, ∞_2, ∞_3 to be diided into four classes, according to the suffixes they carry, and define 'pur' and 'mixed' differences and 'pure' and 'mixed' pairs as he has done.

It is now readily verified that, among the blocks B_1, B_2, B_3, B_4

(i) treatments of the class 1, 2 or 3 occur 9 times, treatments of the class 4, occur ten times, and ∞_1, ∞_2 ∞_3 occur onc. (ii) each pure difference of the type [1, 1], [2, 2] or [3, 3] occurs twice, and each pure difference of the type [4, 4] occurs thrice. (iii) each mixed difference occurs thrice. (iv) each of the classes 1, 2, 3 occurs twice with ∞_1, ∞_2, or ∞_3, whereas the class 4 occurs thrice with ∞_1, ∞_2 or ∞_3.

It therefore follows, just as in the first or second fundamental theorem of Bose, that in the 28 blocks developed from B_1, B_2, B_3, B_4

(a) each treatment belonging to one of the classes 1, 2, 3 is replicated 9 times, whereas each treatment of the class 4 is rplic ted 10 times and ∞_1, ∞_2, ∞_3 are replicated 7 times each. (b) each pure pair of the type [1, 1], [2,2] or [3,3] occurs twice, whereas each pure pair of the type [4,4] occurs thrice. (c) Each mixed pair occurs thrice (d) each pair of which one member is ∞_1, ∞_2 or ∞_3 occurs twice or thrice according as the other member belongs to one of the classes 1, 2, 3 or to the class 4.

We now note that in the adjoined blocks B'_1, B'_2, B'_3

(α) each treatment belonging to one of the classes 1, 2, 3 is replicated once, where ∞_1, ∞_2, ∞_3 are replicated thrice and treatments belonging to the class 4 do not occur; (β) each pure pair of the type [1,1], [2,2], [3,3] occurs once; (γ) each pair of which one member is ∞_1, ∞_2, ∞_3 and the other member belongs to one of the classes 1, 2, 3 occurs once; (δ) each of the pairs $\infty_1\infty_2$, $\infty_2\infty_3$, $\infty_3\infty_1$ occurs thrice.

It follows that in the complete solution consisting of 31 blocks, each reatment is replicated 10 times, and each pair occurs thrice.

4. It is well known that any solution of a symmetrical design (v, b, r, k, λ), leads to the solution of the derived design $(v-k, b-1, r, k-\lambda, \lambda)$, by the process of block section (Bose, 1939), which consists in leaving out one block, and deleting from the other blocks, treatments belonging to this block. From the given solution for the design no. (27), We can thus derive a solution for the design no. (26) with parameters

$$v=21, \quad b=30, \quad r=10, \quad k=7, \quad \lambda=3$$

My thanks are due to Mr. R. C. Bose who suggested the problem and under whose guidance the work was carried out.

REFERENCES

1. BHATTACHARYA, K. N. (1944). A new balanced incomplete Block Design. *Science and Culture* 9(11)
2. BHATTACHARYA, K. N. (1945). On a new Symmetrical Balanced Incomplete Block Design. *Bull. Cal. Math. Soc.*, 36, 91-96.
3. BOSE, R. C. (1939). On the construction of Balanced Incomplete Block Designs. *Ann. Eng.* 9, 358-399
4. ——— (1942) On some new series of Balanced Incomplete Block Designs. *Bull. Cal. Math. Soc.* 34, 17-31.
5. COX, GERTRUDE, M. (1940). Enumeration and construction of Balanced Incomplete Configurations. *Ann. Math. Stat.*, 9, 72-85.
6. FISHER, R. A. AND YATES, F. (1938, 1943). Statistical tables for Biological, Agricultural and Medical Research.
7. HUSSAIN, Q. M. (1945). Symmetrical Incomplete Block Designs with $\lambda=2$, $k=8$ or 9. *Bull. Cal. Math. Soc.*, 37, 115-123.
8. ——— (1946). Impossibility of the symmetrical incomplete block design with $\lambda=2$, $k=7$. *Bull. Cal. Math. Soc.*, 37, 92-94
9. NANDI, H. K. On the relation between certain types of configuration.

Paper received : 25 March 1946.

A STUDY OF EXPENDITURE PATTERNS OF CALCUTTA MIDDLE CLASS FAMILY BUDGETS

By S. JANARDHAN POTI

Statistical Laboratory, Calcutta

The traditional method in family expenditure studies is to start with some classification based on traits fixed arbitrarily from our knowledge of socio-economic conditions such as community, province or some arbitrary economic levels. From such classified data according to immediately noticeable traits attempts are made to calculate average expenditure patterns of each group and to explain the variation of the expenditure patterns due to the basic classification itself or due to causes indirectly involved in the classification itself. In the middle-class budget survey conducted in 1945 by the Indian Statistical Institute data was collected from 1785 family schedules in three independent subsamples. Analysis of the data pertaining to one of the three subsamples on the basis of a classification of families according to total family expenditure reveals that the family expenditure patterns do not appreciably differ from group to group. Table I giving expenditure patterns compiled from data collected from Calcutta diet survey (1945) reveals the truth of the above statement.

TABLE I. EXPENDITURE PATTERNS.

percentage of expenditure on	expenditure groups in rupees.								
	0—50	51—100	101—150	151—200	201—250	251—300	301—400	401—600	600& above
essential food items	60·02	47·56	41·72	39·36	36·30	31·72	31·57	27·17	23·74
non-essential food items	9·00	17·40	20·66	21·08	23·28	25·31	26·14	27·96	28·69
rent & fuel	18·29	16·45	16·45	16·21	14·51	14·51	13·52	13·16	9·51
miscellaneous & clothing	12·69	18·59	21·17	23·35	25·91	28·46	28·77	31·71	38·06
total	100·00	100·00	100·00	100·00	100·00	100·00	100·00	100·00	100·00
average family size	3·27	4·11	4·85	7·14	8·22	8·98	10·36	11·77	16·12

Though the figures show that the patterns consistently vary with total expenditure an examination of the individual family variation reveals that only a small fraction of it is ascribable to total expenditure. There are other causes of variation such as family size, family composition, community, social status etc. which are also equally effective in causing the variation of expenditure patterns. The average family size increases consistently with the expenditure level as shown by the same table. In fact one may state that the classification adopted here does not separate the families into distinct types with respect to their expenditure patterns. Hence any conclusions based on the above type of analysis may not reveal the actual state of affairs. In fact cost of living indices for 1945 for all the groups calculated with 1939 as the base year were found to lie between 280 and 290 though one should expect the cost of living index should be high for high expenditure levels where proportionately more is spent on items (like fish, meat etc.) the prices of which have arisen by 5 or 6 times that of the pre-war price.

Therefore, an attempt has been made in this paper to classify families according to expenditure pattern, and then to investigate further in what respects families possessing the same pattern agree between themselves and in what manner they contrast with families possessing a different pattern. *This method of tackling the problem may help us to classify families into distinct types by means of some common traits.* If no such common traits are found as may happen sometimes when manifold classifications are required to explain the variations of expenditure pattern then we may classify the families according to expenditure pattern itself and study the effect of several factors on each group separately at different times. This is a better method of study of the economic conditions of families at different intervals of time than the one usually adopted which will give a distorted picture of the existing conditions.

The expenditure pattern studied here consists of sets of four variates which are (a) proportion of expenditure on essential food items (by essential food items is meant cereals, pulses, vegetables, spices and gur) to total expenditure (b) proportion of expenditure on non-essential food items (all food items other than included in (a)) to total expenditure (c) proportion of expenditure on rent and fuel to total expenditure (d) proportion of expenditure on miscellaneous and clothing to total expenditure. This effects a classification of all items into four homogeneous classes with price relatives of items in each class close together. As all the items of expenditures are covered up by (a), (b), (c), (d) it is not necessary to consider the four variates as they will add up to one. It is sufficient, therefore, to deal with any three of them only in our statistical analysis.

Since the statistical analysis with a multiplicity of variates is complicated we shall try to search for a function of these variates which can explain a major portion of the variance of the sets of variates. This function we shall call as the most discriminating function and we shall replace the original set of three variates which represented the family pattern by this function.

Each family can be represented by a point in three dimensions with coordinates p_1, p_2, p_3 representing the proportional expenditures on (a), (b) and (c). The variance of such points can be resolved into three components by taking any three mutually orthogonal axes. We shall choose that system of axes for which the sum of squares of the projections of the point vectors on each of the axes is stationary. If the variance of the projections on an axis constitutes a major component of the total variance then the original points may be replaced by their projections on this line. Having obtained the equation of the line we shall get the distance of each projected point from the origin and thus replace the set of variates (p_1, p_2, p_3) by this one variate, namely the distance of the projected point on this line from the origin.

The mathematical derivation of the linear components described above is given below. Let $l_1 p_1+l_2 p_2+l_3 p_3=t$ be one of the components required, then variance of t is

$$\sum_{j=1}^{3} \sum_{i=1}^{3} l_i l_j v_{ij} = Q \qquad (1)$$

where v_{ii} is the variance of p_i and v_{ij} is the covariance between p_i and p_j. Also since the sum of the variances of the three components is equal to the total variances of the set p_1, p_2, p_3 we must have

$$l_1^2+l_2^2+l_3^2-1=v=0 \qquad .. \quad (2)$$

So then our object is to maximise Q with respect to l_1, l_2 and l_3 subject to the restriction (2). Using a Lagrangian multiplier we obtain l_1, l_2 and l_3 as solutions of (2) and (3)

$$\left.\frac{\partial Q}{\partial l_i} - k\,\frac{\partial v}{\partial l_i} = 0 \;\right\} i=1,\,2,\,3 \qquad .. \quad (3)$$

On simplifying (3) we get

$$\left.\begin{array}{l} (V_{11}-k)l_1+V_{12}l_2+V_{13}l_3 = 0 \\ V_{21}l_1+(V_{22}-k)l_2+V_{23}l_3 = 0 \\ V_{31}l_1+V_{32}l_2+(V_{33}-k)l_3 = 0 \end{array}\right\} \qquad .. \quad (4)$$

If solutions other than $l_1=l_2=l_3=0$ exist the determinant of the set of equations (4) must vanish. This gives

$$\begin{vmatrix} (V_{11}-k) & V_{12} & V_{13} \\ V_{21} & (V_{22}-k) & V_{23} \\ V_{31} & V_{32} & (V_{33}-k) \end{vmatrix} = 0 \qquad .. \quad (5)$$

which is a cubic equation in k called the characteristic equation and it is a well known result that such equations have all the roots real and positive. Let k_1, k_2 and k_3 represent the three roots of (5). Substituting any one of them in (4) and multiplying 1st member of (4) by l_1, 2nd by l_2 and 3rd by l_3 and adding up we get $Q-k(l_1^2+l_2^2+l_3^2)=0$ i.e. $Q=k$. Hence each of these roots stand for the variance of each one of the components possessing the properties described above. Substituting each one of the roots k_1, k_2 and k_3 successively in the first two members of (4) and using (2) we determine the corresponding l's by solving in the usual manner. Let us denote the first, second, and third components thus obtained as $l_{i1}p_1+l_{i2}p_2+l_{i3}p_3$, $i=1$, 2 and 3 respectively. It is also easy to see that these components are orthogonal to each other i.e. $\Sigma\, l_{ik}l_{jk}=\delta_{ij}$ where $\delta_{ij}=0$ if $i\neq j$ andone if $i=j$.

One of the three subsamples collected in Calcutta diet survey consisting of 625 family schedules is taken up for analysis. Since 128 family schedules belonged to house owning families whose housing expenses are naturally very low were omitted in our analysis and the results of the analysis on the remaining 497 family schedules are given below. Division of the sums of squares or sums of products by total sample size is avoided to facilitate computational work. The characteristic equation derived from the covariance matrix is

$$\begin{vmatrix} (8{\cdot}4554-k) & -4{\cdot}2233 & {\cdot}0963 \\ -4{\cdot}2233 & (2{\cdot}7997-k) & -1{\cdot}0383 \\ {\cdot}0963 & -1{\cdot}0383 & (1{\cdot}4905-k) \end{vmatrix} = 0$$

which is same as
$$k^3-12{\cdot}7267k^2+21{\cdot}5247k-{\cdot}4018=0$$

The roots of this equation are $k_1=10{\cdot}7461$, $k_2=1{\cdot}9807$, $k_3=0{\cdot}0189$. Total variance is equal to $10{\cdot}7461+1{\cdot}9807+0{\cdot}0189=12{\cdot}7267$. The contribution by the first component to the total variance of the set is 84·4% and that by other two components together is only 15·6% of the total variance. Hence replacing the family pattern (p_1, p_2, p_3) by a single component as the first one we are leaving only about 16% of the total variation cf the expenditure pattern unaccounted and this cannot vitiate the results of our statistical analysis. The three components as well as their variance contributions are given below.

components	direction cosines of the axes l_1	l_2	l_3	contribution to variance
1	·8780	−·4748	·0621	84·4%
2	·2933	·4302	−·8537	15·5%
3	·3785	·7678	·5170	0·1%

Also it can easily be verified that these components are orthogonal to each other.

Taking the first component which explains 84·4% of the total variance to represent the family pattern we find that with increase of family size this component increases but with increase of total expenditure it decreases as is shown by Table 2 giving mean values of the principal component of expenditure pattern by total expenditure groups and family sizes.

TABLE 2. VARIATION OF X (PRINCIPAL COMPONENT) WITH FAMILY SIZE AND TOTAL EXPENDITURE.

family size	expenditure group in rupees 0—50	51—100	101—150	151—200	201—250	251—300	301—400	401—600
2	·41	·30	·10	·02	—	·01	—	·01
3	·52	·30	·16	·08	·06	—	—	—
4	·60	·34	·23	·15	·20	·12	. .	·06
5	—	·39	·29	·21	·26	·16	·03	. .

Except in the case of the groups 151-200 and 201-250 in family sizes 4 and 5 we find that variations are consistent with the statement made above. Such discrepancies as observed are due to samples being too small in those groups.

Let us next study the variation in expenditure pattern due to total expenditure by family sizes. Table 3 below gives the regression equation of x, the principal component on t, the total expenditure for each family size and the contribution of variance due to total expenditure.

TABLE 3. REGRESSION OF THE PRINCIPAL COMPONENT OF EXPENDITURE PATTERN ON TOTAL EXPENDITURE.

family size (1)	no. of families (2)	regression equation (3)	variance of x (4)	variance of x due to t (5)	percentage variation due to t (6)	correlation between x and t (7)
2	34	$(x-{\cdot}2603)=-{\cdot}0013(t-93{\cdot}59)$	·9489	·3715	39·15	−·6257
3	43	$(x-{\cdot}2723)=-{\cdot}0029(t-101{\cdot}67)$	1·4133	·9732	68·86	−·8311
4	70	$(x-{\cdot}2602)=-{\cdot}0011(t-130{\cdot}94)$	1·3842	·5015	26·23	−·6019
5	72	$(x-{\cdot}2992)=-{\cdot}0015(t-131{\cdot}93)$	1·0546	·5575	52·86	−·7271
6	60	$(x-{\cdot}2778)=-{\cdot}0018(t-157{\cdot}96)$	1·2269	·8084	65·89	−·8117
7	46	$(x-{\cdot}2559)=-{\cdot}0010(t-189{\cdot}99)$	·6974	·2867	41·11	−·6412
8	46	$(x-{\cdot}2417)=-{\cdot}0012(t-220{\cdot}47)$	·4311	·2490	57·76	−·7600
9	35	$(x-{\cdot}2699)=-{\cdot}0004(t-271{\cdot}33)$	1·2219	·2169	17·75	−·4213
10	32	$(x-{\cdot}2418)=-{\cdot}0002(t-323{\cdot}72)$	·6330	·1057	16·70	−·4086

It can be seen that even for a given family size the variation brought about by total expenditure is not very appreciable except in the case of families of sizes 3, 5, 6 and 8. If we had analysed the variance irrespective of family size the variance contribution by total expenditure would have been very much smaller as is shown by the following analysis.

Combining families of all sizes we find the correlation between total expenditure and the principal component reduces to −·3290 and the variance contributed by total expenditure to the total variance of the pattern is only 10·82%. This explains why, when classified according to total family expenditure, average family patterns showed little variation between expenditure groups.

Similar analysis has not been done on families whose size exceeds 10 since samples are very small in those cases. The variation from one family size to another cannot be given a functional form (as in the above case we had given a linear fit) since families of a given size belong to a special type distinct from others with respect to composition. As family composition determines to a very larger extent the variation in pattern the variation due to size gets distorted and hence the variation due to family size and total expenditure have not been estimated by applying partial regression formulae to the data.

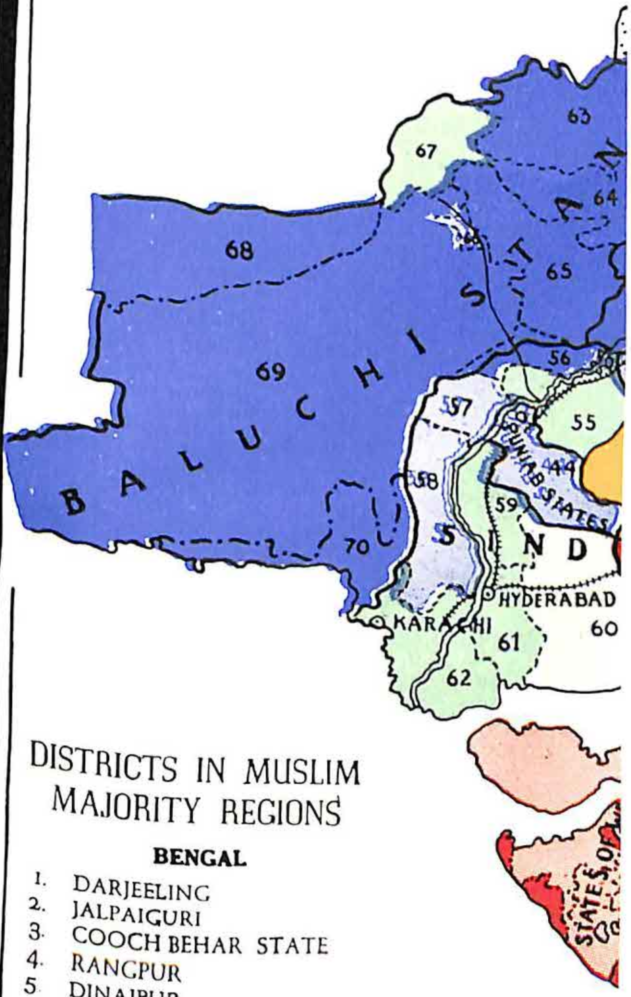
SUMMARY.

(1) Since the principal component $l_{11}p_1+l_{12}p_2+l_{13}p_3$ explains a major portion of the variance, there exists a single economic factor which influences to a great extent the expenditure pattern. This component which increases with family size and decreases with total expenditure can be used to classify families according to poverty levels for budgetary studies.

(2) Though family size and total expenditure together explain in many cases a large part of the variation in expenditure patterns we find that there exist other factors like family composition etc., which too in no small measure determine the poverty level mentioned above.

(3) Since classification by expenditure levels alone cannot divide the families into distinct types with respect to expenditure pattern, any study of changes of pattern of families in a given expenditure level after the lapse of a period leads to erroneous judgement. Therefore, it would be better to group the families according to poverty levels determined by the principal component derived in this article and study how they change their groups and assess improvement or deterioration.

Paper received : 10 May, 1946.



DISTRICTS IN MUSLIM MAJORITY REGIONS

BENGAL

1. DARJEELING
2. JALPAIGURI
3. COOCH BEHAR STATE
4. RANGPUR
5. DINAIPUR
6. MALDA
7. RAJSHAHI
8. BOGRA
9. PABNA
10. MYMENSINGH
11. TRIPURA STATE
12. TIPPERA
13. DACCA
14. FARIDPUR
15. JESSORE
16. NADIA
17. MURSHIDABAD
18. BIRBHUM
19. BURDWAN
20. BANKURA
21. MIDNAPUR
22. HOOGHLY
23. HOWRAH
24. 24-PARGANAS
25. KHULNA
26. BAKARGANJ
27. NOAKHALI
28. CHITTAGONG
29. CHITTAGONG HILL TRACTS

ASSAM

30. SYLHET

PUNJAB

31. ATTOCK
32. RAWALPINDI
33. JHELUM
34. GUJRAT
35. SHAHPUR
36. MIANEALI
37. MUZAFFARGARH
38. JHANG
39.

LYALLPUR

PU

41. GUJRANW
42. SIALKOT
43. GURDASPU
44. PUNJAB ST
45. LAHORE
46. MONTGO
47. MULTAN
48. DERA GAZ

N.

49. DERA ISM
50. BANNU
51. KOHAT
52. PESHAWA
53. MARDAN
54. HAZARA

S

55. SUKKUR
56. UPPER SIN
57. LARKANA
58. DADU
59. NAWABSH
60. THAR PAR
61. HYDERABA
62. KARACHI

BALUC

63. ZHOB
64. LORALAI
65. SIBI
66. BOLAN
67. QUETTA-PI
68. CHAGAI
69. KALAT &

LASBELA

Let us next study the variation in expenditure pattern due to total expenditure by family sizes. Table 3 below gives the regression equation of x, the principal component on t, the total expenditure for each family size and the contribution of variance due to total expenditure.

TABLE 3. REGRESSION OF THE PRINCIPAL COMPONENT OF EXPENDITURE PATTERN ON TOTAL EXPENDITURE.

family size (1)	no. of families (2)	regression equation (3)	variance of x (4)	variance of x due to t (5)	percentage variation due to t (6)	correlation between x and t (7)
2	34	$(x-\cdot 2603)=-\cdot 0013(t-93\cdot 59)$	·9489	·3715	39·15	−·6257
3	43	$(x-\cdot 2723)=-\cdot 0029(t-101\cdot 67)$	1·4133	·9732	68·86	−·8311
4	70	$(x-\cdot 2602)=-\cdot 0011(t-130\cdot 94)$	1·3842	·5015	26·23	−·6019
5	72	$(x-\cdot 2992)=-\cdot 0015(t-131\cdot 93)$	1·0546	·5575	52·86	−·7271
6	60	$(x-\cdot 2778)=-\cdot 0018(t-157\cdot 96)$	1·2269	·8084	65·89	−·8117
7	46	$(x-\cdot 2559)=-\cdot 0010(t-189\cdot 99)$	·6974	·2867	41·11	−·6412
8	46	$(x-\cdot 2417)=-\cdot 0012(t-220\cdot 47)$	·4311	·2490	57·76	−·7600
9	35	$(x-\cdot 2699)=-\cdot 0004(t-271\cdot 33)$	1·2219	·2169	17·75	−·4213
10	32	$(x-\cdot 2418)=-\cdot 0002(t-323\cdot 72)$	·6330	·1057	16·70	−·4086

It can be seen that even for a given family size the variation brought about by total expenditure is not very appreciable except in the case of families of sizes 3, 5, 6 and 8. If we had analysed the variance irrespective of family size the variance contribution by total expenditure would have been very much smaller as is shown by the following analysis.

Combining families of all sizes we find the correlation between total expenditure and the principal component reduces to −·3290 and the variance contributed by total expenditure to the total variance of the pattern is only 10·82%. This explains why, when classified according to total family expenditure, average family patterns showed little variation between expenditure groups.

Similar analysis has not been done on families whose size exceeds 10 since samples are very small in those cases. The variation from one family size to another cannot be given a functional form (as in the above case we had given a linear fit) since families of a given size belong to a special type distinct from others with respect to composition. As family composition determines to a very larger extent the variation in pattern the variation due to size gets distorted and hence the variation due to family size and total expenditure have not been estimated by applying partial regression formulae to the data.

SUMMARY.

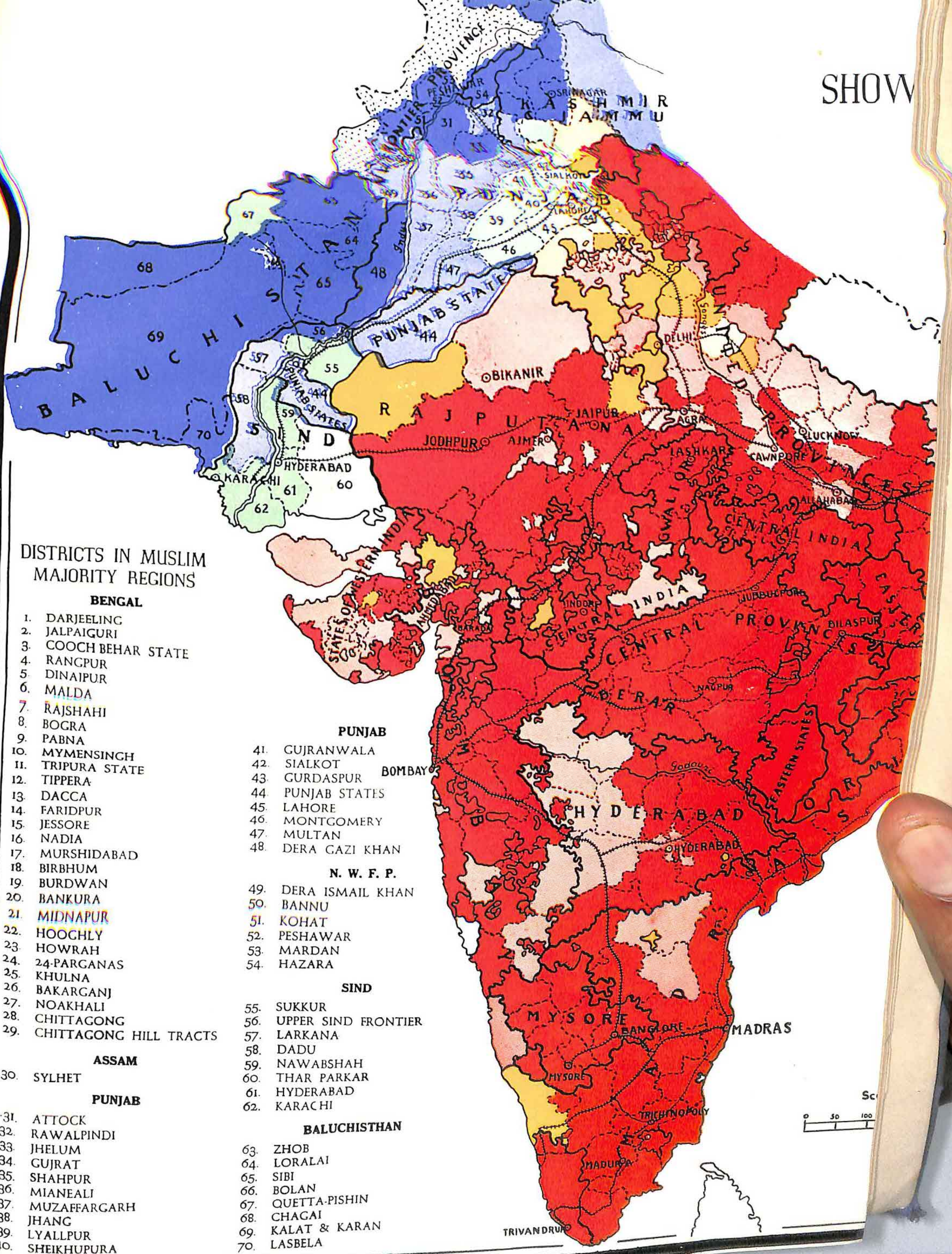
(1) Since the principal component $l_{11}p_1+l_{12}p_2+l_{13}p_3$ explains a major portion of the variance, there exists a single economic factor which influences to a great extent the expenditure pattern. This component which increases with family size and decreases with total expenditure can be used to classify families according to poverty levels for budgetary studies.

(2) Though family size and total expenditure together explain in many cases a large part of the variation in expenditure patterns we find that there exist other factors like family composition etc., which too in no small measure determine the poverty level mentioned above.

(3) Since classification by expenditure levels alone cannot divide the families into distinct types with respect to expenditure pattern, any study of changes of pattern of families in a given expenditure level after the lapse of a period leads to erroneous judgement. Therefore, it would be better to group the families according to poverty levels determined by the principal component derived in this article and study how they change their groups and assess improvement or deterioration.

Paper received : 10 May, 1946.

SHOW



DISTRICTS IN MUSLIM MAJORITY REGIONS

BENGAL

1. DARJEELING
2. JALPAIGURI
3. COOCH BEHAR STATE
4. RANGPUR
5. DINAJPUR
6. MALDA
7. RAJSHAHI
8. BOGRA
9. PABNA
10. MYMENSINGH
11. TRIPURA STATE
12. TIPPERA
13. DACCA
14. FARIDPUR
15. JESSORE
16. NADIA
17. MURSHIDABAD
18. BIRBHUM
19. BURDWAN
20. BANKURA
21. MIDNAPUR
22. HOOGHLY
23. HOWRAH
24. 24-PARGANAS
25. KHULNA
26. BAKARGANJ
27. NOAKHALI
28. CHITTAGONG
29. CHITTAGONG HILL TRACTS

ASSAM

30. SYLHET

PUNJAB

31. ATTOCK
32. RAWALPINDI
33. JHELUM
34. GUJRAT
35. SHAHPUR
36. MIANEALI
37. MUZAFFARGARH
38. JHANG
39. LYALLPUR
40. SHEIKHUPURA

PUNJAB

41. GUJRANWALA
42. SIALKOT
43. GURDASPUR
44. PUNJAB STATES
45. LAHORE
46. MONTGOMERY
47. MULTAN
48. DERA GAZI KHAN

N. W. F. P.

49. DERA ISMAIL KHAN
50. BANNU
51. KOHAT
52. PESHAWAR
53. MARDAN
54. HAZARA

SIND

55. SUKKUR
56. UPPER SIND FRONTIER
57. LARKANA
58. DADU
59. NAWABSHAH
60. THAR PARKAR
61. HYDERABAD
62. KARACHI

BALUCHISTHAN

63. ZHOB
64. LORALAI
65. SIBI
66. BOLAN
67. QUETTA-PISHIN
68. CHAGAI
69. KALAT & KARAN
70. LASBELA

DISTRIBUTION OF MUSLIMS IN THE POPULATION OF INDIA ; 1941.

The accompanying map and tables supply material for a factual study of the distribution of Muslims in India. Eight different shades of colour have been employed to indicate different intensities of Muslims in the population. The whole range (of 100 per cent) has been divided into eight classes each representing an interval of two-anna (or 12½%) of the total population, and one colour has been assigned to each class.

Deep blue represents the highest intensity of Muslims of 14 anna (or 87½% and above,) light blue indicates a Muslim population between 12-anna and 14-anna (between 75% and 87½%), green between 10-anna and 12-anna (62½% and 75%), and light green between 8-anna and 10-anna (between 50% and 62½%) of Muslims.

Orange and red indicate areas in which Muslims form less than 50% of the population. Light orange shows a Muslim proportion between 6-anna and 8-anna (37½% and 50%); orange between 4-anna and 6-anna (25% and 37½%); light red between 2-anna and 4-anna (12½% and 25%) of Muslims ; and finally red represents areas in which Muslims form 2-anna (12½%) or less of the total population.

The tables give by Districts in the different Provinces and States of India, the total population in thousands (col. 2) according to the Indian Census of 1941. The percentage of Muslims is shown in Col. (3) ; the percentage of Hindus in Col. (4) ; and the percentage of 'Others' (which includes tribal groups) in col. (5).

Assam, as a whole, is not a Muslim-majority province, as will be evident from the map itself. The only Muslim-majority region in it is the district of Sylhet which has been numbered and shown in the list of districts given in the lower left-hand side of the map.

Material for the present note was prepared in the Indian Statistical Institute under the supervision of Jitendra Mohan Sen Gupta and N. T. Mathew, and the maps were prepared by the workers of the drawing section among whom Arun Kumar Ganguly, deserves special mention.

P. C. MAHALANOBIS

TABLE SHOWING PERCENTAGE OF MUSLIMS, HINDUS AND OTHERS BY DISTRICTS IN THE DIFFERENT PROVINCES AND STATES IN INDIA IN 1941.

PROVINCES and Districts.	Population in thousands.	Percentage of		
		Muslims	Hindus	Others
(1)	(2)	(3)	(4)	(5)
AJMER-MERWARA				
Ajmer-Merwara	584	15·4	64·4	20·2
ANDAMANS AND NICOBARS				
1. Andamans	21	38·1	38·1	23·8
2. Nicobars	12	0	0	100·0
Total	23	24·2	24·2	51·6
ASSAM				
1. Baliapara Frontier Tract	7	0	42·9	57·1
2. Cachar	641	36·4	35·3	28·3
3. Darrang	737	16·4	47·2	36·4
4. Garo Hills	224	4·9	6·2	88·9
5. Goalpara	1014	46·2	30·2	23·6
6. Kamrup	1264	29·0	55·1	15·9
7. Khasi & Jaintia Hills	119	1·7	10·9	87·4
8. Lakhimpur	895	5·0	56·0	39·0
9. Lushai Hills	153	0	2·0	98·0
10. Naga Hills	190	0·5	2·1	97·4
11. Nowgong	711	35·2	40·6	24·2
12. Sadiya Frontier Tract	60	1·7	30·0	68·3
13. Sib Sagar	1075	4·8	59·8	35·4
14. Sylhet	3117	60·7	36·9	2·4
Total	10207	33·7	41·3	25·0
BALUCHISTHAN				
1. Bolan	6	83·3	16·7	0
2. Chagai	29	96·6	3·4	0
3. Loralai	84	94·0	3·6	2·4
4. Quetta Pishin	156	72·4	18·6	9·0
5. Sibi	165	95·8	3·6	0·6
6. Zhob	61	91·8	6·6	1·6
Total	501	87·6	8·8	3·6
BENGAL				
1. Bakarganj	3549	72·3	27·0	0·7
2. Bankura	1290	4·3	83·6	12·1
3. Birbhum	1048	27·4	65·5	7·1
4. Bogra	1260	84·0	14·8	1·2
5. Burdwan	1891	17·8	73·7	8·5
6. Calcutta	2109	23·6	72·6	3·8
7. Chittagong	2153	74·6	21·3	4·1
8. Chittagong Hill Tracts	247	2·8	2·0	95·2
9. Dacca	4222	67·3	32·2	0·5
10. Darjeeling	376	2·4	47·3	50·3
11. Dinajpur	1927	50·2	40·2	9·6
12. Faridpur	2889	64·8	34·8	0·4
13. Hooghly	1378	15·0	79·8	5·2
14. Howrah	1490	19·9	79·5	0·6
15. Jalpaiguri	1090	23·0	50·6	26·4
16. Jessore	1828	60·2	39·4	0·4
17. Khulna	1943	49·4	50·3	0·3
18. Malda	1233	56·8	37·8	5·4
BENGAL (contd).				
19. Midnapur	3191	7·7	84·1	8·2
20. Murshidabad	1641	56·6	41·7	1·7
21. Mymensingh	6024	77·4	21·5	1·1
22. Nadia	1760	61·2	37·4	1·4
23. Noakhali	2217	81·4	18·6	0·0
24. Pabna	1705	77·1	22·5	0·4
25. Rajshahi	1572	74·6	21·0	4·4
26. Rangpur	2878	71·4	27·9	0·7
27. Tippera	3860	77·1	22·8	0·1
28. 24-Parganas	3536	32·5	65·3	2·2
Total	60307	54·7	41·6	3·7
BIHAR				
1. Bhagalpur	2409	11·6	84·0	4·4
2. Champaran	2398	16·9	82·1	1·0
3. Darbhanga	3457	14·1	85·8	0·1
4. Gaya	2775	10·6	80·1	9·3
5. Hazaribagh	1751	11·9	60·5	27·6
6. Manbhum	2032	6·5	59·7	33·8
7. Monghyr	2565	10·3	87·5	2·2
8. Muzaffarpur	3245	13·1	86·8	0·1
9. Palamau	913	9·3	55·2	35·5
10. Patna	2162	10·9	88·3	0·8
11. Purnea	2390	40·8	54·7	4·5
12. Ranchi	1675	4·7	25·2	70·1
13. Santal Parganas	2234	11·8	37·5	50·7
14. Saran	2861	12·7	86·6	0·7
15. Shahabad	2329	7·6	91·1	1·3
16. Singhbhum	1145	3·8	36·5	59·7
Total	36341	13·0	73·0	14·0
BOMBAY				
1. Ahmedabad	1372	13·0	81·7	5·3
2. Ahmedanagar	1142	5·3	85·2	9·5
3. Belgaum	1225	9·0	85·6	5·4
4. Bijapur	976	12·6	86·9	0·5
5. Bombay City	1490	16·9	68·5	14·6
6. Bombay Suburban	251	11·6	63·7	24·7
7. Broach & Panch Mahals	925	15·6	53·7	30·7
8. Dharwar	1201	14·9	83·1	2·0
9. East Khandesh	1328	11·2	83·2	5·6
10. Kaira	915	9·7	85·7	4·6
11. Kanara	441	7·9	87·5	4·6
12. Kolaba	669	4·9	85·2	9·9
13. Nasik	1114	6·1	77·3	16·6
14. Poona	1359	5·2	88·7	6·1
15. Ratnagiri	1373	6·7	92·6	0·7
16. Satara	1327	3·8	94·4	1·8
17. Sholapur	1015	8·8	89·6	1·6
18. Surat	881	8·6	52·2	39·2
19. Thana	933	4·7	63·2	32·1
20. West Khandesh	912	5·3	54·6	40·1
Total	20849	9·2	79·4	11 4

DISTRIBUTION OF MUSLIMS IN INDIA

TABLE SHOWING PERCENTAGE OF MUSLIMS, HINDUS AND OTHERS BY DISTRICTS IN THE DIFFERENT PROVINCES AND STATES IN INDIA IN 1941.

PROVINCES and Districts.	Population in thousands	Percentage of Muslims	Percentage of Hindus	Percentage of Others
(1)	(2)	(3)	(4)	(5)
CENTRAL PROVINCES & BERAR				
1. Akola	908	11·2	84·2	4·6
2. Amraoti	989	10·0	82·6	7·4
3. Balaghat	634	1·9	75·9	22·2
4. Betul	438	1·8	59·1	39·1
5. Bhandara	963	1·9	86·1	12·0
6. Bilaspur	1549	1·3	79·7	19·0
7. Buldana	821	9·6	87·1	3·3
8. Chanda	873	1·8	78·1	20·1
9. Chhindwara	1034	4·2	57·0	38·8
10. Drug	929	0·9	76·6	22·5
11. Hosangabad	824	4·4	79·7	15·9
12. Jubbulpore	911	7·4	71·9	20·7
13. Mandla	505	1·4	38·0	60·6
14. Nagpur	1060	6·2	86·0	7·8
15. Nimar	513	11·7	64·1	24·2
16. Raipur	1517	1·5	79·8	18·7
17. Saugor	939	4·6	83·4	12·0
18. Wardha	519	4·0	85·4	10·6
19. Yeotmal	888	6·3	75·1	18·6
Total	16814	4·7	76·9	18·4
COORG				
Coorg	169	8·9	77·5	13·6
DELHI				
Delhi	918	33·2	61·8	5·0
MADRAS				
1. Anantapur	1171	10·1	89·2	0·7
2. Bellary	1051	10·8	88·2	1·0
3. Chingliput	1824	2·4	95·0	2·6
4. Chittoor	1632	5·9	93·2	0·9
5. Coimbatore	2810	2·6	94·4	3·0
6. Cuddapah	1057	13·9	81·7	4·4
7. Godavari East	2162	1·5	91·9	6·6
8. Godavari West	1380	2·2	91·5	6·3
9. Guntur	2277	8·0	78·9	13·1
10. Kistna	1444	5·3	85·3	9·4
11. Kurnool	1146	15·1	76·3	8·6
12. Madras	777	12·4	79·9	7·7
13. Madura	2447	4·4	91·7	3·9
14. Malabar	3929	34·0	63·1	2·9
15. Nellore	1617	7·4	87·6	5·0
16. Nilgiris	210	8·6	49·0	42·4
17. North Arcot	2578	6·8	91·2	2·0
18. Ramnad	1980	7·2	87·7	5·1
19. Salem	2869	2·8	96·4	0·8
20. South Arcot	2609	3·2	93·9	2·9
21. South kanara	1524	14·0	72·7	13·3
22. Tanjore	2563	6·6	89·7	3·7
23. Tinnevelly	2245	6·5	81·7	11·8
24 Trichinopoly	2194	4·2	90·7	5·1
25. Vizagapatam	3846	0·8	91·2	8·0
Total	49342	7·9	86·7	5·4
N. W. F. P.				
1. Bannu	296	87·2	10·5	2·3
2. Dera Ismail Khan	298	85·9	13·1	1·0
3. Hazara	796	95·0	3·8	1·2
4. Kohat	289	92·0	5·9	2·1
N. W. F. P. (contd.)				
5. Mardan	507	95·5	2·2	2·3
6. Peshawar	852	90·4	6·0	3·6
Total	3038	91·8	5·9	2·3
ORISSA				
1. Balasore	1029	3·5	93·5	3·0
2. Cuttack	2431	3·0	94·6	2·4
3. Ganjam	1855	0·2	76·1	23·7
4. Koraput	1128	0·3	15·4	84·3
5. Puri	1102	2·4	94·8	2·8
6. Sambalpur	1183	0·4	79·5	20·1
Total	8728	1·7	78·3	20·0
PUNJAB				
1. Ambala	848	31·7	48·4	19·9
2. Amritsar	1414	46·5	15·4	38·1
3. Ahock	676	90·4	6·4	3·2
4. Biloch Trans-Frontier Tract	40	100·0	0·0	0·0
5. Dera Gazikhan	581	88·3	11·5	0·2
6. Ferozepore	1423	45·0	19·6	35·4
7. Gujranwala	912	70·5	11·8	17·7
8. Gujrat	1105	85·5	7·7	6·8
9. Gurdaspur	1154	51·1	24·5	24·4
10. Gurgaon	851	33·6	65·8	0·6
11. Hoshiarpur	1170	32·6	40·0	27·4
12. Hissar	1007	28·3	64·8	6·9
13. Jhang	822	82·6	15·8	1·6
14. Jhelum	630	89·4	6·5	4·1
15. Jullundur	1127	45·2	17·6	37·2
16. Kangra	899	4·8	93·2	2·0
17. Karnal	995	30·6	66·9	2·5
18. Lahore	1695	60·6	16·8	22·6
19. Ludhiana	819	36·9	20·3	42·8
20. Lyallpur	1396	62·9	11·6	25·5
21. Mianwali	506	86·2	12·4	1·4
22. Montgomery	1329	69·2	14·4	16·4
23. Multan	1484	78·0	16·4	5·6
24. Muzaffargarh	713	86·4	12·8	0·8
25. Rowalpindi	785	80·0	10·4	9·6
26. Rahtak	956	17·5	81·6	0·9
27. Shapura	999	83·7	10·1	6·2
28. Sheikhupura	853	63·5	9·1	27·4
29. Sialkot	1190	62·1	19·4	18·5
30. Simla	39	18·0	76·9	5·1
Total	28418	57·1	26·6	16·3
SIND				
1. Dadu	389	84·8	14·9	0·3
2. Hyderabad	759	66·9	32·4	0·7
3. Karachi	714	64·0	31·2	4·8
4. Larkana	511	81·8	17·8	0·4
5. Nawab Shah	584	74·7	24·1	1·2
6. Sukkur	693	71·0	28·1	0·9
7. Thar Parkar	581	50·3	42·7	7·0
8. Upper Sind Frontier	304	90·5	9·5	0·0
Total	4535	70·7	27·1	2·2

TABLE SHOWING PERCENTAGE OF MUSLIMS, HINDUS AND OTHERS BY DISTRICTS IN THE DIFFERENT PROVINCES AND STATES IN INDIA IN 1941.

PROVINCES and Districts	Population in thousands	Percentage.		
		Muslims	Hindus	others
(1)	(2)	(3)	(5)	(5)
UNITED PROVINCE				
1. Agra	1290	15·5	82·6	1·9
2. Aligarh	1373	13·6	83·9	2·5
3. Allahabad	1813	14·3	84·2	1·5
4. Almora	687	0·6	99·1	0·3
5. Azamgarh	1823	12·1	87·7	0·2
6. Bahraich	1241	22·8	77·0	0·2
7. Ballia	1054	6·6	88·3	5·1
8. Banda	723	5·8	94·0	0·2
9. Bara Banki	1163	18·3	81·6	0·1
10. Bareilly	1176	28·4	70·0	1·6
11. Basti	2186	17·8	82·2	0·0
12. Benares	1219	11·2	86·8	2·0
13. Bijnor	910	37·0	61·0	2·0
14. Budaun	1162	18·1	79·4	2·5
15. Buland Shahr	1317	18·5	78·4	3·1
16. Cawnpore	1556	12·9	86·3	0·8
17. Dehra Dun	266	15·4	80·4	4·2
18. Etah	985	11·1	86·3	2·6
19. Etawah	883	6·0	93·2	0·8
20. Farrukhabad	955	12·8	86·7	0·5
21. Fatehpur	807	12·3	87·6	0·1
22. Fyzabad	1319	11·9	88·0	0·1
23. Garhwal	602	0·8	98·8	0·4
24. Ghazipur	985	9·1	88·6	2·3
25. Gonda	1720	19·1	80·9	0·0
26. Gorakhpur	3964	10·8	86·6	2·6
27. Hamirpur	576	6·4	92·0	1·6
28. Hardoi	1239	11·8	88·1	0·1
29. Jalaun	482	6·4	92·3	1·3
30. Jaunpur	1387	8·8	91·0	0·2
31. Jhansi	773	5·6	90·9	3·5
32. Khiri	1024	16·3	83·6	0·1
33. Lucknow	950	23·2	75·5	1·3
34. Mirzapur	900	5·8	89·0	5·2
35. Meerut	1897	23·4	73·4	3·2
36. Mainpuri	873	5·2	93·8	1·0
37. Moradabad	1473	38·5	56·6	4·9
38. Muttra	897	9·8	89·3	0·9
39. Muzaffar Nagar	1057	29·1	68·8	2·1
40. Naini Tal	292	16·1	82·9	1·0
41. Partabgarh	1041	12·3	87·5	0·2
42. Pilibhit	491	21·2	78·2	0·6
43. Rae Barielly	1065	9·6	90·1	0·3
44. Saharanpur	1180	33·6	64·1	2·3
45. Shahjahanpur	983	17 7	81·9	0·4
46. Sitapur	1294	16·4	83·5	0·1
47. Sultanpur	1100	12·6	87·4	0·0
48. Unao	960	9·5	90·5	0·0
Total	55023	15·3	83·2	1·5

STATES	Population in thousands	Percentage.		
		Muslims	Hindus	Others
(1)	(2)	(3)	(4)	(5)
ASSAM				
1. Khasi	214	0·9	11·2	87·9
2. Manipur	512	5·7	59·4	34·9
Total	726	4·3	45·2	50·5
BALUCHISTAN				
1. Kalat	253	96·8	3·2	0·0
2. Karan	34	100·0	0·0	0·0
3. Las Bela	69	97·1	2·9	0·0
Total	356	97·2	2·8	0·0
BARODA				
Baroda	2855	7·8	76·8	15·4
BENGAL				
1. Cooch-Behar	641	37·9	61·6	0·5
2. Mayurbhanj	991	0·6	27·3	72·1
3. Tripura	513	24·0	67·8	8·2
Total.	2145	17·3	47·2	35·5
CENTRAL INDIA				
1. Ajaigarh	97	3·1	85·6	11·3
2. Alirajpur	113	2·6	7·1	9·0
3. Baoni	25	12·0	88·0	0·0
4. Baraondha	17	0·0	82·4	17·6
5. Barwauni	177	5·1	35·0	59·9
6. Bhopaal	785	14·0	76·0	10·0
7. Bijawar i	121	2·5	92·6	4·9
8. Charkher	124	4·8	93·6	1·6
9. Chhatarpur	185	3·8	93·5	2·7
10. Datia	174	4·0	95·4	0·6
11. Dewas (Sr.)	89	10·1	86·5	3·4
12. Dewas (Jn)	84	11·9	81·0	7·1
13. Dhar	253	7·1	54·6	38·3
14. Indore	1514	8·3	71·3	20·4
15. Jaora	117	17·1	76·9	6·0
16. Jhabua	178	1·7	12·9	85·4
17. Khaniadhana	20	5·0	95·0	0·0
18. Khilchipur	49	2·0	93·9	4·1
19. Maihar	80	3·8	76·2	20·0
20. Nagod	88	3·4	85·2	11·4
21. Narsingarh	125	4·0	90·4	5·6
22. Orchha	363	2·5	91·7	5·8
23. Panna	231	2·2	84·0	13·8
24. Rajgarh	149	5·4	91·3	3·3
25. Ratlam	126	11·1	53·2	35·7
26. Rewa	1820	2·1	92·2	5·7
27. Sailana	40	2·5	52·5	45·0
28. Samthar	38	7·9	92·1	0·0
29. Sitaman	33	6·1	87·9	6·0
30. Rest of Central India	296	4·4	76·7	18·9
Total	7511	5·9	78·0	16·1

DISTRIBUTIONS OF MUSLIMS IN INDIA

TABLE SHOWING PERCENTAGE OF MUSLIMS, HINDUS AND OTHERS BY DISTRICTS IN THE DIFFERENT PROVINCES AND STATES IN INDIA IN 1941.

STATES	Population in thousands	Percentage of Muslims		
		Mus-lims	Hindu	Otherss
(1)	(2)	(3)	(4)	(5)
CHHATTISGARH				
1. Bastar	634	0·3	23·7	76·0
2. Changlehakar	21	0·0	33·3	66·7
3. Chhuikhadan	33	3·0	84·8	12·2
4. Jashpur	224	0·9	25·9	73·2
5. Kalahandi	598	0·2	68·1	31·7
6. Konker	149	0·7	34·2	65·1
7. Kawardha	77	1·3	80·5	18·2
8. Khairagarh	174	1·7	85·1	13·2
9. Korea	127	2·4	52·8	44·8
10. Nandgaon	203	1·5	78·3	20·2
11. Panta	632	0·2	78·2	21·6
12. Raigarh	313	0·6	67·7	31·7
13. Sakti	35	1·8	72·7	25·5
14. Sarangarh	141	0·7	99·3	0·0
15. Surguja	552	1·5	24·1	74·4
16. Udaipur	118	0·0	28·0	72·0
Total	4051	0·7	54·0	45·3
COCHIN	1423	7·7	63·0	29·3
DECCAN and KOLHAPUR				
1. Akalkot	104	14·4	84·6	1·0
2. Aundh	89	4·5	94·4	1·1
3. Bhor	156	1·3	94·2	4·5
4. Janjira	104	16·4	80·8	2·8
5. Jath	107	6·5	93·5	0·0
6. Jamkhandi	126	11·1	86·5	2·4
7. Kolhapur	1092	4·5	90·1	5·4
8. Kurundwad (Jn.)	47	10·6	80·9	8·5
9. Kurundwad(Sn.)	53	11·3	81·1	7·6
10. Miraj (Jn)	46	6·5	91·3	2·2
11. Miraj (S.n.)	109	11·9	81·7	6·4
12. Mudhol	72	6·9	93·1	0·0
13. Phaltan	71	2·8	95·8	1·4
14. Ramdurg	40	7·5	92·5	0·0
15. Sangli	293	7·5	87·4	5·1
16. Savanur	22	36·4	63·6	0·0
17. Sawantwadi	252	2·0	95·2	2·8
18. Wadi	2	0·0	100·0	0·0
Total	2785	6·5	89·5	4·0
GUJARAT				
1. Balasiner	61	9·8	88·5	1·7
2. Bansda	55	1·8	14·6	83·6
3. Baria	189	1·6	58·2	40·2
4. Cambay	97	12·4	82·5	5·1
5. Chota Udepur	162	2·5	73·5	24·0
6. Dangs	40	0·0	27·5	72·5
7. Dharampur	123	1·6	18·7	79·7
8. Jawhar	65	1·5	23·1	75·4
9. Lunawada	105	5·7	94·3	0·0
10. Rajpipla	249	3·6	25·3	71·1
11. Sachin	26	11·5	65·4	23·1
12. Sant	94	2·1	38·3	59·6
13. Surgana	18	0·0	94·4	5·6
14. Rest of Agency	173	4·6	64·7	30·7
Total	1457	3·9	52·4	43·7

STATES	Population in thousands	Percentage of		
		Mus-lims	Hindus	Others
(1)	(2)	(3)	(4)	(5)
GWALIOR				
Gwalior	4006	6·0	86·4	7·6
HYDERABAD				
Hyderabad	16339	12·8	81·5	5·7
KASHMIR (including Feudatories)	4022	76·4	20·1	3·5
MADRAS				
1. Banganapalle	45	24·4	71·1	4·5
2. Pudukottai	438	3·9	91·8	4·3
3. Sandur	16	18·8	81·2	0·0
Total	499	6·2	89·6	4·2
MYSORE				
Mysore	7329	6·6	91·2	2·2
ORISSA				
1. Athgarh	55	0·0	87·3	12·7
2. Athmallik	73	0·0	79·4	20·6
3. Bamra	178	0·0	58·4	41·6
4. Baramba	53	0·0	90·6	9·4
5. Baudh	146	0·0	84·2	15·8
6. Bonai	93	0·0	38·7	61·3
7. Daspalla	54	0·0	79·6	20·4
8. Dhenkanal	324	0·3	85·8	13·9
9. Gangpur	398	1·3	23·9	74·8
10. Hindol	58	0·0	94·8	5·2
11. Keonjhar	530	0·4	71·1	28·5
12. Khandpara	87	0·0	95·4	4·6
13. Kharsawan	51	2·0	39·2	58·8
14. Narsinghpur	48	0·0	95·8	4·2
15. Nayagarh	161	0·6	94·4	5·0
16. Nilgiri	73	0·0	82·2	17·8
17. Pal Lahara	34	0·0	67·6	32·4
18. Rairakhol	38	0·0	76·3	23·7
19. Ranpur	51	0·0	94·1	5·9
20. Seraikela	155	1·3	46·4	52·3
21. Sonepur	249	0·4	94·0	5·6
22. Talchar	86	0·0	94·2	5·8
23. Tigiria	26	3·8	92·3	3·9
Total	3021	0·5	70·7	28·8
PUNJAB				
1. Bahawalpur	1341	82·0	12·6	5·4
2. Chamba	169	7·1	92·3	0·6
3. Dujana	31	22·6	77·4	0·0
4. Faridkot	199	30·6	11·1	58·3
5. Jind	362	14·1	74·0	11·9
6. Kapurthala	378	56·6	16·1	27·3
7. Khairpur	306	83·0	16·3	0·7
8. Loharu	28	14·3	85·7	0·0
9. Malirkotla	88	38·6	26·1	35·3
10. Mandi	233	1·7	97·8	0·5
11. Nabha	340	26·6	43·2	36·2
12. Pataudi	22	15·2	81·5	0·0
13. Patiala	1936	22·6	30·8	46·6
14. Suket	71	1·4	98·6	0·0
Total	5504	40·9	33·7	25·4

TABLE SHOWING PERCENTAGE OF MUSLIMS, HINDUS AND OTHERS BY DISTRICTS IN THE DIFFERENT PROVINCES AND STATES IN INDIA IN 1941.

STATES	Population in thousands	Percentage		
		Muslims	Hindus	others
(1)	(2)	(3)	(4)	(5)
PUNJAB HILL				
1. Bilaspur	110	1·8	98·2	0·0
2. Kalsia	67	37·3	44·8	17·9
3. Sirmoor	156	4·5	93·6	1·9
4. Tehri Garhwal	397	0·5	99·5	0·0
5. Other States	360	3·1	96·1	0·8
Total	1090	4·3	94·0	1·7
RAJPUTANA				
1. Alwar	823	26·7	65·6	7·7
2. Banswara	259	2·7	29·0	68·3
3. Bharatpur	576	19·3	78·0	2·7
4. Bikaner	1293	14·3	76·8	8·9
5. Bundi	249	4·8	74·7	20·5
6. Danta	31	9·7	51·6	38·7
7. Dholpur	287	6·6	88·2	5·2
8. Dungarpur	274	3·3	37·2	59·5
9. Jaipur	3041	8·2	81·0	10·8
10. Jaisalmer	93	29·0	67·7	3·3
11. Jhalawar	122	8·2	86·1	5·7
12. Karauli	152	5·3	71·7	23·0
13. Kishengarh	104	6·7	88·5	4·8
14. Kotah	777	6·8	79·0	14·2
15. Kushalgarh Chief Ship	41	2·4	12·2	84·4
1. Lawa (Estate)	3	0·0	100·0	0·0
2. Marwar	2556	8·4	83·1	8·5
3. Mewar	1927	3·3	69·3	27·4
4. Patanpur	316	9·8	81·0	9·2
5. Partabgarh	92	4·4	46·7	48·9
6. Shahpura	61	4·9	85·2	9·9
7. Sirohit Abu	239	3·4	70·3	26·3
8. Tonk	354	14·7	76·8	8·5
Total	13670	9·5	75·5	15·0
SIKKIM				
Sikkim	122	0·0	37·7	62·3

STATES	Population in thousands	Percentage.		
		Muslims	Hindus	others
(1)	(2)	(3)	(4)	(5)
TRAVANCORE				
Travancore	6070	7·2	58·4	34·4
U.P.				
1. Benares	451	8·4	90·7	0·9
2. Rampur	477	49·3	49·9	0·8
Total	928	29·4	69·7	0·9
WESTERN INDIA				
1. Bhavnagar	618	8·1	87·7	4·2
2. Cutch	501	23·4	63·5	13·1
3. Dhangadhra	94	5·3	90·4	4·3
4. Dhrol	34	11·8	85·3	2·9
5. Eastern Kathiawar Agency	323	6·8	87·9	5·3
6. Gondal	245	16·3	80·4	3·3
7. Idar	308	3·6	94·8	1·6
8. Jafrabad	14	14·3	85·7	0·0
9. Junagadh	671	19·1	79·6	1·3
10. Limbdi	44	11·4	84·1	4·5
11. Morvi	142	9·2	85·9	4·9
12. Nawanagar	504	13·3	81·4	5·3
13. Palitana	76	6·6	86·8	6·6
14. Porbandar	147	9·5	89·1	1·4
15. Radhanpur	68	13·2	82·4	4·4
16. Rajkot	103	10·7	82·5	6·8
17. Sabar Kanthar Agency	458	3·9	91·5	4·6
18. Vijayanagar	14	0·0	35·7	64·3
19. Wadhwan	51	5·9	82·4	1·7
20. Wankaner	55	29·1	65·4	5·5
21. Western kathiawar Agency	436	14·2	83·7	2·1
Total	4906	12·3	82·9	4·8

MISCELLANEOUS NOTES

TEST OF SIGNIFICANCE FOR INTRA-CLASS CORRELATION WHEN FAMILY SIZES ARE NOT EQUAL.

By R. P. BHARGAVA

Statistical Laboratory, Calcutta.

INTRODUCTION.

The test of significance for the intra-class correlation when family sizes are equal was given by Fisher in his book 'Statistical Methods for Research Workers".

In practice, we often meet with cases where family sizes are unequal. Thus, if we consider a certain character among brothers in a family, it often happens that the number of brothers in the different families is not the same. In such cases, some of the problems that arise are,

(1) test of significance, that is, whether population correlation coefficient is equal to any assigned value ρ

(2) to find the fiducial limits for ρ

(3) to determine the best estimate of population correlation ρ.

(4) to get a suitable estimate of population mean m.

(5) to estimate intra-class correlation from family means of two unequal family sizes, the individual observations of the family being not known.

2. TESTS OF SIGNIFICANCE.

Let us suppose there are n families and let the k_i individuals of the ith family give the observations

$$x_{i1},\ x_{i2},\ldots x_{i}k_i\ (i = 1,\ 2,\ldots n)$$

We consider $x_{i1}, x_{i2}, \ldots x_i k_i$ to be a random observation from a k_i variate normal population in which each of the variates has the same mean m, and the same standard deviation σ, the correlation between any two variates being equal to ρ. It may be noted that as m, σ and ρ are not functions of i, each family has the same mean m, the same standard deviation σ and the same correlation coefficient ρ.

Let $\bar{x}_i$ be the sample mean of the ith family. The probability of occurrence of the observation in the i-th family is given by

$$\text{Const.}\ e^{-\frac{1}{2\sigma^2(1-\rho)}\left[\sum_{j=1}^{k_i}(x_{ij}-m)^2-\frac{\rho}{1+\overline{k_i-1}\,\rho}\sum_{j,\,j'=1}^{k_i}(x_{ij}-m)(x_ij'-m)\right]}\prod_{j=1}^{k_i}dx_{ij}$$

where $\text{Const.}=(\sqrt{2\pi}\sigma)^{-k_i}(1-\rho)^{-(k_i-1)/2}(1+\overline{k_i-1}\,\rho)^{-1/2}$

This can be written as

$$\text{Const}\ .\ e^{-\frac{1}{2\sigma^2(1-\rho)}\left[\sum_{j=1}^{k_i}(x_{ij}-\bar{x}_i)^2+\frac{(1-\rho)\ k_i}{1+\overline{k_i-1}\,\rho}(\bar{x}_i-m)^2\right]}\prod_{j=1}^{k_i}dx_i; \tag{2·1}$$

Hence the joint distribution of the sets of observations of our sample can be written as

$$\text{Const.}\ e^{-\frac{1}{2\sigma^2(1-\rho)}\left[\sum_{i,j}(x_{ij}-\bar{x}_i)^2+(1-\rho)\ \Sigma\frac{k_i}{1+\overline{k_j-1}\,\rho}(\bar{x}_i-m)^2\right]}\Pi dx_{ij} \tag{2·2}$$

(2·3)

Let

$$c^2{}_i=\frac{k_i(1-\rho)}{1+\overline{k_i-1}\,\rho},\ \Sigma c^2{}_i\bar{x}_i=\Sigma c^3{}_i\bar{x}$$

Then (2.2) can be written as

$$\text{Const. } e^{-\frac{1}{2\sigma^2(1-\rho)}\left[\Sigma(x_{ij}-\bar{x}_i)^2+\Sigma c_i^2(\bar{x}_i-\bar{x})^2+\Sigma c_i^2(\bar{x}-m)^2\right]} \quad \Pi dx_{ij} \tag{2·4}$$

From this it is easily seen that

$$F=\frac{\Sigma c^2_i(\bar{x}_i-\bar{x})^2}{(n-1)} \cdot \frac{\Sigma(k_i-1)}{\Sigma(x_{ij}-\bar{x}_i)^2} \tag{2·5}$$

is distributed in the usual F distribution with $n_1= n-1$ and $n_2=(\Sigma k_i-1)$ degrees of freedom.

This F statistic can be used for testing whether our families (sizes being unequal) came from multivariate normal populations having population correlation ρ. Thus as a particular case if we desire to know, whether our sample of n families with unequal sizes came from multivariate normal populations having $\rho=·5$, all that we have to do is, to calculate c^2_i and $\bar{x}$ by putting $\rho=·5$, and then calculate the value of F by using (2·5) and consult the F tables.

Thus we can with the help of the F distribution answer any of the following questions (*i*) whether population correlation is significantly different from ·5 (say), (*ii*) whether the population correlation coefficient is significantly greater than ·5 and (*iii*) whether the population correlation coefficient is significantly less than ·5.

These questions are the same as those which arise in testing the significance of intra-class correlation for families of equal size. But it is useful to note that for answering question (*i*) we have to consider both the tail-ends of the F distribution, while for question (*ii*) we have only to consider the tail-end on the right and for (*iii*) the tail-end on the left.

It is of interest to note that (*i*) Fisher's Z-test for testing the significance of intra-class correlation for equal family size and (*ii*) Fisher's technique of analysis of variance for testing the significance of intra-class correlation when family sizes may not be equal follow also from this test.

(*i*) Thus if in (2·5), we put $k_1=k_2='\ldots=k_n=k$ say then F becomes

$$\begin{aligned} F&=\frac{1-\rho}{1+\overline{k-1}\,\rho}\;\frac{\Sigma k(\bar{x}_i-\bar{x})^2}{n-1}\;\frac{n(k-1)}{\Sigma(x_{ij}-\bar{x}_i)^2} \\ &=\frac{1-\rho}{1+\overline{k-1}\,\rho}\;\frac{1+\overline{k-1}\,r}{1-r}\;\frac{n}{n-1} \end{aligned} \tag{2.7}$$

where $\bar{x}$ is the grand mean and r is the sample intra-class correlation. And it is easily seen that Fisher's Z is the same as $\frac{1}{2}\log_e F$.

(*ii*) Again putting $\rho=o$, F is equal to

$$F=\frac{\Sigma k_i(\bar{x}_i-\bar{x})^2}{n-1}\;\frac{\Sigma(k_i-1)}{\Sigma(x_{ij}-\bar{x}_i)^2}$$

which is the same as that used for testing the significance of between and within variation in analysis of variance.

3. FIDUCIAL LIMITS OF ρ

Next we consider the problem of ascertaining fiducial limits of population ρ for the case of unequal family sizes. The usual procedure for finding the lower fiducial limit is to first find (say) the 5 % point of F corresponding to $n_1= n-1$, $n_2= \Sigma(k_i-1)$. Let us call this point as F_5 then from the equality

$$\frac{\Sigma c^2_i(\bar{x}_i-\bar{x})^2}{n-1}\;\frac{\Sigma(k_i-1)}{\Sigma(x_{ij}-\bar{x}_i)^2}=F_5$$

we solve for ρ. From this we get that $\rho>\rho_o$ in 95 % cases. This provides us with the lower limit of ρ. But in our case, the solution does not come out very easily, hence we shall follow the method of trial and error and graduation. Let us calculate values of F of (2.5) for various values of ρ say $\rho=$ .1, .2, .3, .4, .5, ,6, .7, .8, .9.

This can be presented in a tabular form as follows :—

ρ	$F(\rho)$		$F_5\ (n_1, n_2)$
$\cdot1$	$1 < F(\rho = \cdot1)$	$>$	$F_5\ (n_1 = n-1,\ n_2 = \Sigma(k_i - 1)$
$\cdot2$	$1 < F(\rho = \cdot2)$	$>$	$F_5\ (n_1 = n-1,\ n_2 = \Sigma(k_i - 1)$
$\cdot3$	$1 < F(\rho = \cdot3)$	$<$	$F_5\ (n_1 = n-1,\ n_2 = \Sigma(k_i - 1)$
$\cdot4$	$1 = F(\rho = \cdot4)$		
$\cdot5$	$1 > F(\rho = \cdot5)$	$>$	$\frac{1}{F_5}\ (n_1 = \Sigma k_i - 1,\ n_2 = n-1)$
$\cdot6$	$1 > F(\rho = \cdot6)$	$>$	$\frac{1}{F_5}\ (n_1 = \Sigma k_i - 1,\ n_2 = n-1)$
$\cdot7$	$1 > F(\rho = \cdot7)$	$<$	$\frac{1}{F_5}\ (n_1 = \Sigma k_i - 1,\ n_2 = n-1)$
$\cdot8$	$1 > F(\rho = \cdot8)$	$<$	$\frac{1}{F_5}\ (n_1 = \Sigma k_i - 1,\ n_2 = n-1)$
$\cdot9$	$1 > F(\rho = \cdot9)$	$<$	$\frac{1}{F_5}\ (n_1 = \Sigma k_i - 1,\ n_2 = n-1)$

In the above table $F_5 = (n_1, n_2)$ in column (3) indicate the 5 % point of F table corresponding to n_1 and n_2 degrees of freedom. Column (2) gives the value of F corresponding to particular values of ρ.

In the above hypothetical table it is clear that the fiducial limits of ρ are given by

$$\cdot2 < \rho < \cdot7$$

This gives a rough idea, and some more numerical vlues of ρ between .2 and .3, and .7 and .6 will (with the help of graduation or otherwise) give closer approximation for the fiducial limits.

4. Best estimate of ρ

Next comes the problem of obtaining the best estimate of ρ. The value of ρ in the above table which makes F equal to unity is the best estimate of ρ, in the sense of maximising the likelihood of $z = \frac{1}{2}\log_e F$.

Thus from the above table we see that $\rho = .4$ gives $F = 1$. Thus our point estimate of ρ from the given sample comes as .4. This may be called unbiassed sample intra-class correlation r.

It is clear from the way we have proceeded that our method of determining fiducial limits and point estimate of ρ ; coincide with the usual methods when family sizes are equal.

5. Estimate of m

From (4) it is clear that the distribution of $\bar{x}$ is

$$\text{Const}\ .\ e^{-\frac{1}{2\sigma^2}\Sigma\frac{k_i}{1+\overline{k_i-1}\,\rho}(\bar{x}-m)^2}\,d\bar{x} \tag{5·1}$$

where

$$\bar{x} = \frac{\Sigma\frac{k_i x_i}{1+\overline{k_i-1}\,\rho}}{\Sigma\frac{k_i}{1+\overline{k_i-1}\,\rho}}$$

Thus $\bar{x}$ is a sufficient statistic to estimate m. Thus if population ρ is known, we can calculate the best estimate of m. But, as is usual, if ρ is not known, we can estimate m, substituting for ρ, the sample intra-class correlation r, and this gives us a consistent estimate for m. Thus

$$\bar{x} = \frac{\Sigma\frac{k_i\bar{x}_i}{1+\overline{k_i-1}\,r}}{\Sigma\frac{k_i}{1+\overline{k_i-1}\,r}}$$

is a consistent sample estimate of m. The use of this very often arises in crop surveys, budget enquiries etc

6. ESTIMATION OF INTRA-CLASS CORRELATIONS FROM TWO SETS OF FAMILIES

Let us suppose that there are n families of size k and n' families of size k', ($k'<k$) coming respectively from k and k' variate normal populations having the same mean m, the same standard deviation σ, and the same correlation ρ. The individual readings in a family are not known but the family means (or totals) are known. The problem is to get an estimate of ρ.

Let $\bar{x}_i$ be the mean of the ith family of size k ; and let $\bar{x}'_{i'}$ be the mean of the i'th family of size k', Then

$$n\bar{x}=\sum_{i=1}^{n} x_i, \text{ and } n'\bar{x}' = \sum_{i'=1}^{n'} \bar{x}'_{i'} \tag{6·1}$$

also

$$(n-1)s^2_1=\Sigma \mathrm{k}(\bar{x}_i-\bar{x})^2, \text{ and } (n'-1)s^2_2=\Sigma \mathrm{k}'(\bar{x}_{i'}'-\bar{x}')^2 \tag{6·2}$$

Then from (2.4), it is easily seen that

$$F=\frac{1+(\mathrm{k}'-1)\,\rho}{1+(\mathrm{k}-1)\,\rho}\;\frac{s^2_1}{s^2_2} \tag{6·3}$$

is distributed in the usual F distribution with $n_1=n-1$ and $n_2=n'-1$ degrees of freedom.

From this, very often we can get an estimate of ρ. The numerical value of ρ, lying between $\frac{-1}{\mathrm{k}-1} < \rho < 1$ which makes F of (6.3) equal to unity, is the best estimate of ρ from family means.

It is further seen that whenever, the ratio of the two mean squares s_1^2/s_2^2 is $>\mathrm{k}/\mathrm{k}'$ no such numerical value of ρ lying in $\frac{-1}{\mathrm{k}-1} < \rho < 1$ exists which will make F of (6.3) equal to unity : but on the other hand, as is usually the case, if the ratio s_1^2/s_2^2 is $<\mathrm{k}/\mathrm{k}'$ such a numerical value of ρ always exists. And in such cases this estimate is given by

$$\mathrm{r}=-1/(\mathrm{k}-1)+\frac{s^2_1}{s^2_1-s^2_2}(\mathrm{k}-\mathrm{k}')$$

SUMMARY.

(1) We have given a test of significance (unknown) for testing whether our sample of n families, family size being unequal, has come from multivariate normal populations having population correlation coefficient equal to ρ.

(2) We have given a method of deriving the fiducial limits of ρ.

(3) We have given a new definition for intra-class correlation when family size is unequal. This definition reduces to the usual definition for the equal family size and is the best estimate of ρ.

(4) We have derived a consistent estimate for the population mean m.

(5) We have discussed a connected problem of estimation of intra-class correlaton from family means of two unequal family sizes when individual readings of the family are not known but only family means or totals are known.

REFERENCES.

(1) FISHER, R. A. (1944) Statistical Methods for Research workers, 9th Edition.
(2) ——————(1938) Statistical Theory of Estimation, Calcutta University Readership Lectures.
(3) KENDALL, M. G. (1943) The Advanced Theory of Statistics, Vol. I.

Paper received : 4 April 1946.

ON LOCALLY MOST POWERFUL TESTS WHEN ALTERNATIVES ARE ONE SIDED

By C. RADHAKRISHNA RAO and S. JANARDHAN POTI

Statistical Laboratory, Calcutta

The concept of a most powerful test with respect to a single admissable alternative to a specified hypothesis H_0 is due to Neyman and Pearson who, in an extensive series of articles, have advanced a general theory of testing of hypotheses. A test is said to be uniformly most powerful if it is most powerful with respect to every admissable alternative. The rarity of such tests lead Neyman and Pearson to consider tests giving maximum power for alternatives in the neighbourhood of the specified hypothesis (the hypothesis being assumed measurable). Following their ideas we propose to give here tests which are locally powerful with respect to alternatives on one side only. This problem arose in connection with certain investigations that are being carried on by the authors where the altrnatives on one side alone are taken into account.

Denoting the probability density of a set of variables by $p(x, \theta)$ and the specified value of the parameter θ by θ_0, we define a region w of the sample space as locally most powerful for one-sided alternatives if

$$(i) \quad \int_w p\,(x, \theta_0)\ \pi dx = \alpha \text{ (level of significance)}$$

$$(ii) \quad c \int_w p'\,(x, \theta_0)\ \pi dx \quad \text{is maximum}$$

are satisfied where c is + 1 or − 1 according as the alternatives are $\theta > \theta_0$ or $\theta < \theta_0$. The following theorem gives the method of determining such regions.

Theorem : *If whatever may be the region w in the sample space the integral $\int_w p'\,(x, \theta_0)\ \pi dx$ exists, then the region w_0 within which $cp'\,(x, \theta_0) \geqslant \lambda p\,(x, \theta_0)$ and outside which $cp'\,(x, \theta_0) \leqslant \lambda p(x, \theta_0)$ where λ is so determined that $\int_w p\, x, \theta_0) = \alpha$ is locally most powerful for alternatives $\theta > \theta_0$* and *$\theta < \theta_0$ according as c is + 1 or − 1.*

The proof is a direct consequence of a lemma due to Neyman and Pearson (on page 11 in *Research Memoirs* Vol 1, (1936). The boundary of the above region consists of the surfaces of constant values of $p'(x, \theta_0) \,/\, p\,(x, \theta_0) = \phi'\,(x, \theta_0))$ where ϕ is the *likelihood* defined by Fisher. We need only find the distribution of ϕ and/or tabulate values such that the probability of $\phi'\,(x, \theta_0)$ (*i*) exceeding them is α for alternatives on one side and (*ii*) exceeded by them is α for alternatives on the other side. On the other hand if the surfaces of constant ϕ' coincide with surfaces of constant values of a statistic T, then tabulated values of T may be used in tests of significance.

An important appliction of this result when the size of the sample is large emerges from the following theorem.

Theorem : *If (i) $x_1, x_2, \ldots x_n$ form* n *independent observations from a probability distribution, (ii) the information (as defined by Fisher) of a single observation is finite and indicated by I, then, to test the hypothesis that the value of a parameter θ, occurring in the probability distribution, is θ_0 the statistic $\phi'\,(x, \theta_0)/\sqrt{In}$ where ϕ is the liklihood of the sample, can for sufficiently large n be used as a normal deviate and in judging significance only one tail of the normal curve has to be considered when the alternatives are one-sided.*

The proof follows from the fact that $\phi\,(x, \theta_0)$ is approximately normally distributed due to central limit thereom and $E\,[\,(x, \theta_0)] = o$ and $V\,[\,(x, \theta)] = n\,I$.

Paper received : 20 June 1946

ON THE MOST EFFICIENT DESIGNS IN WEIGHING

By C. RADHAKRISHNA RAO

Statistical Laboratory, Calcutta

There are p objects to be weighed and n weighments can be made. In case there is a bias in the scales this may be considered as an additional weight which is used in all weighments. The observational equations are of the form $Y=TA'$, where $Y=(y_1\, y_2 \ldots y_n)$ is the row matrix of recorded weights, $T=(\tau_1, \tau_2, \ldots \tau_p)$ is the row matrix of weights to be determined and A is a matrix with n rows and p columns with elements of the form 0, +1 or −1 (0, +1, and −1 correspond to the cases where the object is not used, or is put in the first pan or second pan). If we require that all weights including bias, if any, should be determined with equal accuracy and that this should be the maximum possible, it follows that the matirx A should satisfy the equation $A'A=nI$ if the dispersion matrix of Y is of the form $\sigma^2 I$. This equation is satisfied if the elements of A are either +1 or −1 and any two of its rows are orthogonal.

We may take the first row to be (1, 1, ...1). If the second row contains k_1 elements with +1 and k_2 with −1, then the orthogonality condition reduces to $k_1=k_2=k$ whence n is of the form $2k$. If a third row exists which is orthogonal to the first two then we can by similar algebra show that n is of the form 2^2k and in this case one more row could be constructed by multiplying the corresponding elements in the second and third rows so that the four rows are mutually orthogonal. By introducing variables giving the number of positives and negatives common to a newly introduced row and the other rows and solving the system of equations due to orthogonality we arrive at the following solution. *If* $n=2^s(2m+1)$, *then there can be at the most* 2^s *rows with elements of the form* +1 *or* −1, *which are mutually orthogonal to one another*. When n is a full power of 2 we have n such rows. The above method gives a procedure for getting at the matrix A′ in all cases. The number of objects p including bias, for the best possible design under consideration is less than or equal to 2^s when the number of weighments n is of the form $2^s(2m+1)$. In this case the weight of each object including bias is estimable with the variance σ^2/n and the estimate of σ^2 can be based on $(n-p)$ degrees of freedom obtained from the sum of squares of residuals of the normal equations. Some designs of special type have been considered by Hottelling (1944) and Kishen (1945).

It is interesting to note that in the case the total weight of the objects is known (a contingency which may arise when an object of known weight breaks into peices and their wieghts are to be determined) the weighment corresponding to the first row need not be made while others are necessary for the best possible design. The methods of deriving the normal equations, variances and covariances of estimates in the case of observational equations with restrictions on the parameters have been discussed by the author in (Rao : 1945).

Another important observation is that all the combinations of weights may not have the same error in weighing. This however requires experimental verification.

REFERENCES

1. HOTELLING, H. (1944) On efficient designs in weighing. *Ann. of Math. Stat.*, 15, 297-306.
2. KISHEN, K. (1945) On the design of experiments in weighing and other experimental technique. *Ann. of Math. Stat.* 16, 295-300.
3. RAO, C. R. (1945) Markoff's theorem with linear restrictions on parameters. *Sankhyā* **7**, Part 1. 16-19.

Paper received : 10 July 1945

A NOTE ON CONCORDANCE

By MONI MOHAN MUKHERJEE
Statistical Laboratory, Calcutta.

If there are k items of work paid at the rate of c_1 per hour for the i-th work, usually the rates are fixed on subjective estimation of the hardness of the work. Sometimes, an objective estimation of the rates is also made by setting up normal outputs. That is, number of units of a particular work which can be done by a normal worker is obtained experimentally and a unit is paid less when a large number of units can be finished in an hour. But here also the value of the hour is determined *a priori*. In this note I suggest that the money equivalent of the hour may be fixed with reference to the preference of the individual workers.

Let the ranks of the k items of work according to the first worker be $a_{11}, a_{21}, a_{31}, \ldots a_{k1}$ where a's are permutations of the integers 1, 2, . k. Ultimately we get the rank matrix as

$$\begin{vmatrix} a_{11} & a_{12} & \cdot\cdot & \cdot\cdot & a_{1m} \\ a_{21} & a_{22} & \cdot\cdot & \cdot\cdot & a_{2m} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ a_{k1} & a_{k2} & \cdot\cdot & \cdot\cdot & a_{km} \end{vmatrix} \qquad \ldots (1)$$

Provided there is concordance between the rankings of the different workers, it will be logical to make the money equivalent of the hour depend on $\bar{a}_i = \frac{1}{m}\sum_{j=1}^{m} a_{ij}$. That is we will adopt the principle of paying less for the most preferred work.

For example if $\bar{a}_1 > \bar{a}_2 > \cdots > \bar{a}_k$ we can make $c_1 > c_2 > \cdots > c_k$.

Given a rank matrix in the above form, divergence from a purely random structure can be tested by χ^2 test where

$$\chi^2 = \frac{12m}{k(k+1)}\sum_{i=1}^{k}\left(\bar{a}_i - \frac{k+1}{2}\right)^2 \qquad \ldots (2)$$

In an actual experiment concerning 10 items of work of 87 workers the value of χ^2 was found to be as large as 184·88 for 9 degrees of freedom. This clearly shows that the pattern in the matrix is by no means random.

Co-efficient of concordance is defined as $w = 12s/m^2(k^3-k)$ where $s = m^2\Sigma(\bar{a}_i-\bar{a})^2$, $\bar{a}$ being the general mean. The value of the coefficient lies between 0 and 1. If ρ_{av} is the average value of the Spearman's rank correlation between any two possible pairs of columns in (1), then it can be shown that $\rho_{av} = (mw-1)/(m-1)$. In our experiment the value of w has come out to be 0.2493 which in itself does not indicate high concordance.

A rigorous test of significance of w has not been arrived at for large values of m and k, but in such case, $z = \frac{1}{2}\log_e[(m-1)w/(1-w)]$ with $\nu_1 = (k-1)-2/m$ and $\nu_2 = (m-1)[k-1)-2/m]$ degrees of freedom can be shown to follow the z distribution. Our value of z comes to be 2.3519 with 9 and 772 degrees of freedom. From z tables this is clearly seen to be significant. This conclusively proves that the rank matrix as a whole shows concordance.

The actual mean values $\bar{a}_i$ are given below :

$\bar{a}_1$	$\bar{a}_2$	$\bar{a}_3$	$\bar{a}_4$	$\bar{a}_5$	$\bar{a}_6$	$\bar{a}_7$	$\bar{a}_8$	$\bar{a}_9$	$\bar{a}_{10}$
8·3	6·9	3·1	5·9	5.5	3.8	5·2	4·3	5·8	6·0

In continuation of what has been stated earlier it may be pointed out that the money equivalent of the hour can be made to depend on these values. Some method of testing significance of the difference $\bar{a}_p - \bar{a}_q$ will add immensely to the value of the method. As it is, there is very little force in the assertion that the ninth work is more preferred than the fourth.

Paper received : 31 January 1946.

NOTE ON CONCENTRATION OF AGRICULTURAL WEALTH IN BENGAL

BY RAMKRISHNA MUKHERJEA,
Statistical Laboratory, Calcutta,.

The author (1946) jointly with Mani Mohon Mukherjee has recently shown that in two samples of village people in Bengal the income distributions have been found to be sharply differentiated and polarised. In the samples from Birbhum and Bogra respectively, 47 % and 41 % of the total income had been found to be concentrated in the hands of only 10% of all families, while the remaining 90% of families had only 53 % and 59 %, that is rest of the income. The samples were not large enough to draw final conclusions but the close agreement between distributions relating to two sets of villages lying wide apart (one sample from West and the other from North Bengal) probably indicates a general tendency.

Bengal is an agrarian country of which about 75 % of the population depend on agriculture. In consequence, such a concentration of income is obviously the result of marked inequality in land distribution. Paddy is the main crop of the province being sown on about 88 % of all cultivated land in Bengal. In a recent study of the economic condition of rural Bengal (*A sample survey of after-effects of the Bengal Famine of* 1943 by P. C. Mahalanobis, Ramkrishna Mukherjea and Ambika Ghosh), it has been shown that in April 1943, before the famine, 36 % of all rural families did not possess any paddy land. 41% had only upto 2 acres, 15 % had between 2 and 5 acres, and only 8 % had above 5 acres of paddy land. This well defined hierarchy and acute concentration of paddy land in the hands of a few families explain the polarisation of income in rural areas to a great extent.

Such polarisation of income arising out of property inequality is further enhanced by the similar concentration of agricultural wealth. This can be realised by studying the different types of people interested in land in relation to production and the proportional distribution of these types in the society as a whole. As noted earlier, paddy is the main crop of the province, so that the production relation of the different types as regards the cultivation of this crop will indicate the nature and extent of concentration of agricultural wealth. Also it is only in the cultivation of paddy (because it is the main crop) all the types of interest are found to occur, such as, self-cultivation, cultivation by hired labour (fourd in regard to all other crops) as well as share-cropping.

There are in all seven types of interest in regard to the cultivaton of paddy.

(1) State, receiving the revenue from the land.

(2) *Zamindar* and subinfeudatory landlords who collect the revenue and along with it realisea share of the agricultural wealth as rent.

(3) *Jotdar* who lends his holding for share-cropping and receives half share of the crop in return of paying the rent only.

(4) Rich farmer who owns the land as a non-cultivating owner and have it cultivated by the aricultural labourers.

(5) *Ryot*, the traditional peasant, who owns the land and cultivates it himself with occasional help of an agricultural labourer in times of expedieney, such as, for transplantation of *aman* paddy and harvest. He gets the total crop.

(6) *Bargadar* who cultivates the land lent by the *jotdar* and receives half share of the crop, bearing the entire cost of production minus the rent of the land.

(7) Agricultural labourer who serves as the hired labour for the rich farmer (and occasionally for a *ryot*) and derives his income as daily wages.

Aman , *aus* and *boro* are the three varieties of paddy grown in Bengal. Of these *boro* is grown on a negligible area. *Aus* comprises of about 23% and *aman* about 77% of the total area under paddy. The land under *aus* is usually never kept fallow for the rest of the year, after the harvest of the crop, and *aman* is grown on that field. Therefore, the study of production relation in regard to the *aman* paddy alone (the principal food and money crop of the province) will indicate the nature of concentration of agricultural wealth in Bengal.

The Table 1 below gives the amount and proportion of neti ncome to gross obtained from an acre of and under *aman* paddy by the seven types of interest as noted above. The table is derived from table V (*b*) and appendix II of the "*Report on the cost of production of crops in he principal sugar and cotton tracts in India, volume VI Bengal*" published by the Imperial Council of Agricultural Research in 1938. The survey was carried out in six villages of Rajshahi-Bogra and six villages of Birbhum districts by regular observation for a period of three years, 1934-35, 1935-36 and 1936-37. The authenticity of the data, therefore cannot be doubted.

In the table the income of the agricultural labourer has been derived from the proportion of the cost of production spent on human labour. The income of *bargadar* has been derived as half share of the total crop plus the interest earned on the capital invested for cultivation (cattle, plough, etc.) minus the cost of production from which rent has been deducted. The income of the *ryot* is the farm business income derived as the margin of profit from the cash price and the cost of production of the crop plus the interest earned on the capital invested. The income of the rich farmer is derived as the farm business income minus the cost on hired labour. The *jotdar's* income is half the share of produce minus the rent. The *zamindar's* income is the rent minus the revenue to the State which is the income of the latter. Incidence of revenue in the two are as has been adopted from "*Man behind the plough*" by Sir Azizul Haque.

It should be noted that the cost of production does not include the charges for self-labour.

Table 1. Average net and gross income received from an acre of land under *aman* paddy by the interested types during 1934-35, 1935-36 and 1936-37.

interested types	net income in rupees		percentage of net income to gross.	
	Rajshahi & Bogra.	Birbhum	Rajshahi & Bogra.	Birbhum.
(1)	(2)	(3)	(4)	(5)
State	0·6—0·5	1·0	2	3
Zamindar	1·3—1·4	3·1	4	9
Jotdar	13·8	13·6	44	38
Rich farmer	13·2	11·3	42	32
Ryot	19·8	14·8	63	42
Bargadar	8·3	5·5	26	15
Agricultural labourer	8:9	7·8	28	22
Gross income	31·4	35·5	100	100

It can be seen from the table that in both the areas, Rajshahi & Bogra and Birbhum, the proportion of net income to gross is highest for the *jotdar* (barring the *ryot* for the obvious reason that he owns the land and does the whole job himself) and lowest for the *bargadar*. The income of the latter is even lower than that of the agricultural labourer. The income of rich farmer is close but comes next to that of the *jotdar* even though he has to spend some time in supervision over the hired labourers as well as for the maintenance and provision of plough cattle, plough, seed, manure and such other necessaries of cultivation. The *jotdar*, on the other hand, has nothing to bother about. Incidentally it is worth mentioning that the income of the *zamindar* and the subinfeudatory landlords is found to be two to three times the revenue to the State indicating how terrible is the exploitation of our peasantry from various sources.

Similar maldistribution of agricultural wealth has also been noted in a survey of six villages in Bogra by the author in 1942 as shown in Table 2 below. It may be noted that well defined income hierarchy and marked concentration of income had been found in this sample of rural society as referred to earlier.

Table 2. Net and gross income received from an acre of land under *aman* paddy by the interested types during 1941-42 in the sample of six villages from Bogra.

interested types	net income (in rupees)	percentage of net to gross income.
(1)	(2)	(3)
State	0·5	1
Zemindar	4	9
Jotdar	18	40
Rich farmer	16	36
Ryot	31	69
Bargadar	15	32
Agricultural labour	16	36
Gross income	45	100

The tables indicate the nature of concentration of agricultural wealth in Bengal. Its extent, however, as is obvious from the tables, depends on the extent of share-cropping and cultivation by hired labourer in contrast to self-cultivation. The Land Revenue Commission Bengal 1940, expressed the view that "5 acres would be the minimum area required to keep an average family in reasonble comfort and if the land is capable of growing nothing but *aman* paddy the area required would be about 8 acres." Therefore, the agricultural population having less than 5 acres of paddy land is bound to depend on share-cropping for an economic holding or have to take to the occupation of agricultural labourer. On the other hand, self-cultivation can only be practised on economic holdings, that is, by the families owning atleast "above 5" acres of paddy land. But obviously in this group also belong the rich farmer and the *jotdar*. The extent of share-cropping and cultivation by hired labour can now be realised from the fact that, as noted earlier, 92 % of all rural families possess paddy land below 5 acres and it is only for the remaining 8% that the holdings are of the size of above 5 acres.

Census of 1901 and 1941 indicate that the population of agricultural labourer has increased from 4·86 millions to 6·34 millions during the last 40 years. Unfortunately, precise statistical data on the number of share-cropers is hardly available. According to the previously mentioned Land Revenue Commission " two-fifths of the agricultural families", that is 40%, have taken to share-cropping, while 29% are agricultural labourers. This is quite possible because, as indicated by the tables, the relationship of the *jotdar* and the *bargadar* in our rural economy gives the highest profit to the non-cultivating owner and so the existence of the disintegrated peasentry (with uneconomic holdings as a result of loss of land) is conditioned by it.

Numerically the *iotdar* and the rich farmer represent a very small minority as they comprise of only a section of 8 % of all rural families in Bengal. But the concentration of income in this group due to the concentration of the *means of production*, (namely land, as referred to earlier,) is further enhanced by the concentration of agricultural wealth through share-cropping and cultivation by hired labour. In consequence, this group is undoubtedly the master of rural Bengal today because of its importance in peasant economy as a whole, both in regard to the share it has of the total *means of production* owned by the peasantry and to its share of the total produce produced by the peasantry.

The author proposes to discuss the inter-relationship in the rural economy arising out of such concentration of agricultural wealth and income in a future communication.

REFERENCE

MUKHERJEA, R AND MUKERJEE, M.M. (1946). A note on concentration of income in Bengal villages. *Sankhya* 7 (3)., p. 327-28

Paper received : 20 March 1946.

A STATISTICAL STUDY OF LABOUR IN THE ASSAM TEA PLANTATION*

By M. V. SESHAGIRI RAO

Labour Department, Simla.

For Assam Plantations statistics of emigration, employment, wages, mortality etc. are available from the year 1930 onwards from the annual reports of the Controller of Emigrant Labour. The object of this paper is to study the trends in emigration, employment, wages and mortality as reflected in the statistics available.

Table I gives the annual figures of *emigration* for the period 1930-44, of men, women and children to the Assam tea gardens. The period under consideration car be divided broadly into three smaller periods; 1930-34, the years of the great depression, 1935-39, the period of recovery and 1939-44, the period of the second World War. The year 1929 saw the onset of the great depression which affected the agricultural countries more adversely than industrial countries, though the latter also suffered serious dislocation, and the effect on the tea industry can easily be noticed in the figures of emigration during this period. The number of emigrants has steadily decreased during the 5 or 6 years from its initial value of about 60 thousands to about 20 thousands in 1935—an immense decrease of about 65 percent. Although a slight improvement was noticed since the year 1934 as is reflected by the year to year increase in emigration, recruitment never reached the predepression levels. The increase in recruitment during the years 1935-36 inspite of the unsatisfactory economic condition of the tea industry, was partly explained by the bad harvests and scarcity that prevailed in the recruiting areas and a consequent labour supply many times more than what was actually required wherefrom the employer could make a choice as to the kind of labour that would be useful. The general conditions of the tea industry were definitely better during the years 1937 and 1938. Again from 1939, when war broke out, there was a considerable decrease in the volume of emigration till the year 1942. The reasons for this fall seem to be the following: (i) Certain districts like Goalpara, Kamrup and Cachar began to suspend recruitment from outside and employ the native labour. (ii) Recruitment was strictly confined to married couples and all illegal cases were severely prosecuted. This resulted in a great fall in the recruitment of men as compared with women and children. (iii) There was a regular demand for labour everywhere for military purposes and the labourers were not very anxious to migrate to the tea gardens on the comparatively very low wages. The year 1943 has however shown a large increase in the volume of emigration, about three times that in the previous year. This was mainly due probably to the employment of the Indian Tea Association labour in the defence projects of Assam.

TABLE I. ANNUAL FIGURES OF EMIGRATION TO THE ASSAM TEA GARDENS

Year	men	women	children	total
1930	33,510	14,117	12,169	59,796
1931	30,245	13,362	9,912	53,519
1932	23,247	14,986	12,764	50,997
1933	15,412	12,552	11,937	39,900
1934	17,483	14,617	15,860	47,960
1935	7,317	6,286	6,365	19,968
1936	8,436	7,587	7,856	23,879
1937	10,286	8,979	8,577	27,842
1938	12,645	10,729	9,641	33,015
1939	11,014	9,933	9,315	30,262
1940	9,488	7,792	7,497	24,777
1941	6,835	6,095	6,066	18,996
1942	5,512	4,810	4,249	14,511
1943	19,267	19,198	19,575	58,040
1944	16,457	12,925	13,558	41,940

---

* Being printed with the kind permission of the Government of India Labour Department.

We may next consider the changes if any in the working of the tea gardens from year to year. These figures are evidently different from the figures of emigration in as much as they include the emigrants already on the plantations and also the local labour, which in fact forms a major part in the working in some places. It is of importance to us as reflecting to a certain extent the volume of production and the general economy of the industry as a whole. Table 2 gives the average working strength of the estates and the figures include the settled as well as the *faltu* or *basti* labour. The estimated average working strength in each year was based on the figures for two representative months in the year viz., March and September and the other months were not taken into consideration. It will be seen from the table that the average working strength, unlike the corresponding figures cf emigration, has remained almost stationary during the period varying narrowly about the mean value 464,065 with a standard deviation of 17,044. The dispersion from the mean nowhere exceeds twice the standard deviation. Thus, although the number of emigrants has been repidly decreasing year by year, the working strength of the tea estates has been stationary except for a few minor fluctuations which are insignificant. The fall in the figures of emigration therefore indicates under the present conditions that the demand for emigrant labour has reached the saturation level.

TABLE 2. AVERAGE DAILY NUMBER OF WORKERS SETTLED AND FALTU

year	men	women	children	total
1934	216,468	180,527	54,790	451,785
1935	218,742	165,133	51,720	435,595
1936	230,653	185,906	59,522	476,081
1937	220,697	178,596	58,487	457,780
1938	218,409	181,197	62,644	462,250
1939	227,281	191,999	64,918	484,198
1940	231,084	189,407	68,425	488,916
1941	214,205	185,841	67,122	467,168
1942	221,423	180,629	75,252	477,302
1943	206,033	186,021	76,183	468,237
1944	189,498	173,849	72,053	435,400

Another interesting fact that can be noticed from the two tables considered above, is the definite increase in the proportion of female and child labour as against the decided decrease in that of the costlier male labour. Thus, whereas the proportion of men, women and children recruited in 1930 was 17:7:6, the same in 1943 stands at 1:1:1. This is no doubt a direct consequence of the restriction on emigrant labour, confining recruitment to married couples alone. Though not so striking as in the case of recruitment, the same tendency can be noticed to a certain extent in the working strength also.

TABLE 3. AVERAGE DAILY EARNINGS OF SETTLED LABOURERS ON THE TEA ESTATES

Year	Darrang	Goalpara	Kamrup	Lakhimpur	Nowgong	Sibsagar	Cachar	Sylhet	Average
	as. ps.	as. ps.	as. ps.	as. ps.	as. ps.	as. ps.	as. ps.	as. ps.	as. ps.
1934	5 2	2 3	4 5	5 7	4 2	4 11	3 11	2 5	4 4
1935	4 9	2 0	3 3	6 1	4 8	3 8	4 5	2 6	4 2
1936	4 7	5 0	4 5	5 11	4 8	3 8	4 0	2 8	4 3
1937	4 11	5 1	3 7	6 1	4 8	3 10	4 0	2 7	4 3
1938	4 11	5 1	4 5	6 1	5 1	3 2	4 2	2 9	4 4
1939	5 1	5 2	4 6	6 7	5 2	3 9	4 3	3 2	4 8
1940	5 8	3 1	4 10	6 5	5 3	3 10	4 3	4 5	4 11
1941	5 9	3 3	5 5	6 8	5 3	4 7	5 10	4 10	5 4
1942	5 10	6 1	5 8	7 0	5 11	6 0	5 2	4 9	5 9
1943	5 0	6 5	—	7 3	5 10	6 3	5 9	4 10	5 11
1944	6 11	7 7	7 2	8 2	6 10	6 11	5 5	5 4	6 8

Let us now consider the question of wages. The Committee which inquired into labour conditions in Assam in 1922 found that on some gardens wages had been unchanged for almost a quarter of a century. Even the planters themselves, it seems, came to recognise by 1918 that a rise was imperatively necessary and granted it in 1920. It is not known what the position was afterwards till the early thirtees, as detailed statistics are not available for this period. Table 3 gives the weighted average daily earnings of settled labourers comprising of men, women and children. These figures also, it may be pointed out, were calculated on the basis of the figures for two representative months in every year viz., March and September. The simple averages of the figures for the several districts are given in the last column of Table 3. This and the following calculations exclude the figures for the District Goalpara as they are found to be too irregular to be considered reliable: and are confined to the period 1934-41 the latter years being abnormal due to the outbreak of war. The first thing that is to be noticed from these figures is the very low level of the wages which clearly gives an idea of the deplorable living standard of the garden coolies. The planters have noticed this fact long back and have found it easier to retain the labourers attached to their estates by giving them pieces of agricultural land thus 'converting them as part-time agriculturists. It is also said that the workers are provided with certain facilities like free fuel supply and cheap grain supply particularly during the period of inflation. While there is a considerable difference between the wages paid in the several districts of Assam—the highest in Lakhimpur and the lowest in Sylhet—there was practically very little movement in the wage levels in each district till the outbreak of the last war. To determine the exact trend of movement, a straight line is fitted up to the average earnings for the years 1934-41 by the method of least squares, which gives

$$Y = 1.7X + 46.9$$

where Y gives the average earnings in pies and X stands for the years that have elapsed since 1933.

According to the linear movement assumed, the average yearly rate of increase in the daily earnings is 1.7 pies. Thus, the average daily earnings have, according to the graduated values, risen from Rs. -4-1 a day in 1934 to Rs. -4-9 in 1939 and Rs. -5- in 1941. The years following have witnessed a considerable increase in the actual average earnings reaching Rs. -6-8 in 1944, and this is perhaps mainly due to dearness allowances paid in cash, which came into existence all over India after 1942. But this increase, when compared with the unprecedented rise in the cost of living, is very insignificant. Summing up the results of the above analysis, we observe that there has been paractically no rise in the wage level throughout the first half of the decade under consideration ; and it is doubtful if there is any during the latter half. The apparent rise during the later half may as well be attributed to the extra allowances given, which did not keep pace with the rising cost of living and one wonders if the actual economic position of the garden labourer has not become worse towards the end of the period under cosideration. It may be pointed out here that the average cash earnings of the *faltu* or *basti* labourers are much lower than those of the settled labourers considered above.

Let us now consider how far the climate and atmosphere of the tea gardens have been affecting the emigrant labour. Table 4 gives the number of deaths per thousand among the tea garden labourers as compared with the death rates in general for the province of Assam as given in the Statistical Abstract.

TABLE 4. COMPARISON OF THE DEATH RATES OF GARDEN LABOURERS WITH THR GENERAL DEATH RATES

year	death rates per thousand		difference observed
	in gardens	in Assam	
1934	22·6	19·0	+3·6
1935	22·6	20·5	+2·1
1936	22·8	18·6	+4·2
1937	22·9	20·9	+2·0
1938	23·1	20·9	+2·2

In all the given years considered above, the death rates in the tea gardens are observed to be far higher than these in the Province in general.

TABLE 5. PERCENTAGES OF DEATHS DUE TO THE SEVERAL DISEASES

year	Respiratory	Malaria	Dysentry	Diarrhoea	Anaemia
1934	16·7	11·0	9·6	6·0	6·5
1935	19·0	12·5	13·0	6·5	6·4
1936	19·7	12·5	11·5	6·6	6·0
1937	18·8	12·3	11·3	6·9	6·0
1938	20·5	11·0	11·6	6·6	6·1
1939	18·6	11·8	11·6	7·6	5·8
1940	18·0	10·4	9·5	6·6	5·6
1941	14·8	11·9	9·4	7·2	6·4

Table 5 gives the percentages of deaths due to the several important diseases in the plantations. Deaths from the respiratory deseases appear to be the most frequent, as about 15 to 20 percent of the deaths every year belong to this class. Malaria and Dysentry account for about 10 to 12 percent of the deaths each. Next in frequency are Diarrhoea and Anaemia. Thus these five types are found to result in more than 50 percent of the fatalities. The high frequency of the respiratory diseases gives place for doubt regarding the suitability of the garden climate particularly to the emigrant labour. This is a matter on which first hand investigaton is necessary if we are to form any conclusive opinions.

In conclusion, we may connect the various results arrived at as follows. It is certain that the volume of emigration has been very rapidly declining year by year. The increase in 1943 over the previous year was due to special reasons which are temporary and the very next year namely 1944 has witnessed a considerable fall. The Indian Tea Association labourers who have been taken up in the defence projects of Assam are to return back now that the war is over, and turn to their original occupation. That is likely to result in a further fall in emigration in the years to come. Turning to the original trend of events, we find that there are two possible reasons for the downward tendency in the volume of emigration, suggesting themselves. Firstly, it is just possible that the planters themselves are tending towards an economic method of employing more and more of native labour in place of imported labours which is more costly. Secondly, the labourers themselves are perhaps not very much willing to migrate to a far off place where the prospects and conditions are no better than in their own places if not worse. The replacement of male labour by the cheaper female and child labour is a testimony to the apparent tendency of the planters to be economical. A study of the figures of wages and mortality has clearly shown that the conditions in the plantations are not quite satisfactory. It may therefore be concluded that the two possible reasons put forth for the decrease in the figures of emigration year by year are not mutually exclusive, but have been working simultaneously to produce it. If in addition, due to the expected depression, prices fall and production is to be curtailed to a large extent, the volume of emigration which has recently risen to its initial heights, may again experience a rapid and drastic decline.

Paper received 11 February, 1946

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

The Hony. Secretary, Indian Statistical Institute, wishes to record his acknowledgment with thanks of the donation of Rs. 1000/-received from the Rockefeller Foundation in two instalments through the National Institute of Sciences of India which has made possible the publication of the following papers which could not have been published otherwise :—

1. ON THE LINEAR COMBINATION OF OBSERVATIONS AND THE GENERAL THEORY OF LEAST SQUARES. By C. Radhakrishna Rao. *Sankhya, Vol.* 7(3), 1946.
2. CONFOUNDED FACTORIAL DESIGNS IN QUASI-LATIN SQUARES. By C. Radhakrishna Rao. *Sankhya, Vol.* 7 (3), 1946.
3. ENUMERATION OF NON-ISOMORPHIC SOLUTIONS OF BALANCED INCOMPLETE BLOCK DESIGNS. By Hari Kinkar Nandi. *Sankhya, Vol.* 7 (3), 1946.
4. A FURTHER NOTE ON NON-ISOMORPHIC SOLUTIONS OF INCOMPLETE BLOCK DESIGNS. By Hari Kinkar Nandi. *Sankhya, Vol.* 7 (3), 1946.
5. ON THE ORDER OF APPROXIMATION INVOLVED IN LAPLACE'S CENTRAL LIMIT THEOREM IN PROBABILITY. By Manindra Nath Ghosh *Sankhya, Vol.* 7 (3), 1946.
6. A NOTE ON CONCENTRATION OF INCOME IN BENGAL VILLAGES. By Ramkrishna Mukherjea and Moni Mohan Mukherjee. *Sankhya, Vol.* 7 (3), 1946.
7. ON A MEASURE OF DIVERGENCE OF TWO MULTINOMIAL POPULATIONS. By A. Bhattacharyya. *Sankhya, Vol.* 7 (4), 1946.
8. DISCRIMINANT FUNCTIONS IN MULTIVARIATE ANALYSIS. By C. R. Rao. *Sankhya, Vol.* 7 (4), 1946.
9. A STUDY OF EXPENDITURE PATTERN OF CALCUTTA MIDDLE CLASS FAMILY BUDGETS. By S. J. Poti. *Sankhya, Vol.* 7 (4), 1946.
10. ON THE MOST EFFICIENT DESIGNS IN WEIGHING. By C. R. Rao. *Sankhya Vol.* 7 (4), 1946.